

181 TAACAACGAGTAATTAGCAAAATCCGAGCAGAAAACCTCACCACCACTCCGAAATTCACG 240

Db 181 TAACAACGAGTAATTAGCAAAATCCGAGCAGAAAACCTCTCACCCACCTCCGAAATTCACG 240
Qy 241 TCCTTCACTAAATTTTCGAAAGGAATCGATCAATACCAACCCCAATACACAAAATACATAA 300
Db 241 TCCTTCACTAAATTTTCGAAAGGAATCGATCAATACCAACCCCAATACACAAAATACATAA 300
Qy 301 TCAAAATGGCGAGAAATCGTACCTGGAAAATTGCTTCAAGTCGAGAGAGAGGAAAGGA 360
Db 301 TCAAAATGGCGAGAAATCGTACCTGGAAAATTGCTTCAAGTCGAGAGAGAGGAAAGGA 360
Qy 361 AGATCGTGGAGAAAGGGTTTAGGGTTTAAAGCTCAGACTTCTATTGGAGTAAATGGGACG 420
Db 361 AGATCGTGGAGAAAGGGTTTAGGGTTTAAAGCTCAGACTTCTATTGGAGTAAATGGGACG 420
Qy 421 GTGTACATTTTCCGTTTGGAAAATGAATCTTTGGGCTCAAGTTATGGGCTATTAGATATT 480
Db 421 GTGTACATTTTCCGTTTGGAAAATGAATCTTTGGGCTCAAGTTATGGGCTATTAGATATT 480
Qy 481 TGATGGGCTTCTAGTAAATACAAATATAGTTATTTGGGCTTAGTTTAAATGAAGCCCATGT 540
Db 481 TGATGGGCTTCTAGTAAATACAAATATAGTTATTTGGGCTTAGTTTAAATGAAGCCCATGT 540
Qy 541 TGGAAATATTTGACACATGCTTGGCTTACTAGTCTAAACATGCAACCGAACAGTTGTGCG 600
Db 541 TGGAAATATTTGACACATGCTTGGCTTACTAGTCTAAACATGCAACCGAACAGTTGTGCG 600
Qy 601 AGACAAGTCGAGCATATACAAATGGAATGGAATGGAATGGAATGGAATGGAATGGAATGGA 660
Db 601 AGACAAGTCGAGCATATACAAATGGAATGGAATGGAATGGAATGGAATGGAATGGAATGGA 660
Qy 661 GTGTACCTTGTCTGCTGCTTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAAT 720
Db 661 GTGTACCTTGTCTGCTGCTTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAAT 720
Qy 721 ATACAAATTTTGGCTGATCTTGCACAACTCTCGATCATATCGCAATATACGTGAACA 780
Db 721 ATACAAATTTTGGCTGATCTTGCACAACTCTCGATCATATCGCAATATACGTGAACA 780
Qy 781 CTGGTGATCTAATTTGTTGTTGTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAAT 840
Db 781 CTGGTGATCTAATTTGTTGTTGTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAAT 840
Qy 841 TGAATATATGATCATGTTTAAACATTTGTAAGTAAATGATATTAATGATTAATTTTAAATTT 900
Db 841 TGAATATATGATCATGTTTAAACATTTGTAAGTAAATGATATTAATGATTAATTTTAAATTT 900
Qy 901 AGTTGATGATAACGTAAGCAAAATGAGATAGATACATTTGATTTTGTGCTATTTTG 960
Db 901 AGTTGATGATAACGTAAGCAAAATGAGATAGATACATTTGATTTTGTGCTATTTTG 960
Qy 961 ACATATGCGGAGTGAGCTACGCGCATGAAATCAAGAGACACTTGTCTGAGCTCACAG 1020
Db 961 ACATATGCGGAGTGAGCTACGCGCATGAAATCAAGAGACACTTGTCTGAGCTCACAG 1020
Qy 1021 AGTACGCTGTAAGAGCTTAGACTCAAGTCCCATGCAACCTAATCTCTAGTCTGCTCA 1080
Db 1021 AGTACGCTGTAAGAGCTTAGACTCAAGTCCCATGCAACCTAATCTCTAGTCTGCTCA 1080
Qy 1081 ACCAGAGCTCACTTGACAAATATATAAACTCTCTAGTCTGCTGCTCTCTCTCTCTCTCTCT 1140
Db 1081 ACCAGAGCTCACTTGACAAATATATAAACTCTCTAGTCTGCTGCTCTCTCTCTCTCTCTCT 1140
Qy 1141 CTCACAAACAAACAAAAG 1158
Db 1141 CTCACAAACAAACAAAAG 1158

RESULT 2
AX707061
LOCUS AX707061 1164 bp DNA linear PAT 04-APR-2003
DEFINITION Sequence 13 from Patent WO03014347.
ACCESSION AX707061
VERSION AX707061.1 GI:29563372

KEYWORDS

SOURCE Arabidopsis thaliana (thale cress)

ORGANISM

Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis;

REFERENCE

Ohlrogge, J.B., Benning, C., Gao, H., Girke, T.A. and White, J.A.

Plant seed specific promoters

TITLE Patent: WO 03014347-A 13 20-FEB-2003;

JOURNAL

MICHIGAN STATE UNIVERSITY (US)

FEATURES

Location/Qualifiers

1..1164

source

/organism="Arabidopsis thaliana"

/mol_type="unassigned DNA"

/db_xref="taxon:3702"

ORIGIN

Query Match 100.0%; Score 1158; DB 6; Length 1164;

Best Local Similarity 100.0%; Pred. No. 4.3e-228;

Matches 1158; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CACAAACATACACTCAAAATCCAGACTCAGATCTCAATATATGCAACTTCATCATGAA 60

Db 1 CACAAACATACACTCAAAATCCAGACTCAGATCTCAATATATGCAACTTCATCATGAA 60

Qy 61 AACATCAAAACAGTCAAAATCAAGTCAAGTTCAGCAGACCAAGCCAGTAAA 120

Db 61 AACATCAAAACAGTCAAAATCAAGTCAAGTTCAGCAGACCAAGCCAGTAAA 120

Qy 121 GATAGAAAATTTAAGCAACGCTCATGCTAAGCTCGCAAAATATCTTCTAATCAAAACAG 180

Db 121 GATAGAAAATTTAAGCAACGCTCATGCTAAGCTCGCAAAATATCTTCTAATCAAAACAG 180

Qy 181 TAACAACGAGTAAATTTAGCAAAATCCGAGCAGAAAACTCTACCACCTTCGAAATTCACG 240

Db 181 TAACAACGAGTAAATTTAGCAAAATCCGAGCAGAAAACTCTACCACCTTCGAAATTCACG 240

Qy 241 TCCTTCACTAAATTTTGGAAAGGAATCGATCAATACCAACCCATTACACAAAATACATAA 300

Db 241 TCCTTCACTAAATTTTGGAAAGGAATCGATCAATACCAACCCATTACACAAAATACATAA 300

Qy 301 TCAAAATGGCGAGAAATCGTACCTGCAAACTTTGCTTCAAGTCGAGAGAGAGGAAAGGA 360

Db 301 TCAAAATGGCGAGAAATCGTACCTGCAAACTTTGCTTCAAGTCGAGAGAGAGGAAAGGA 360

Qy 361 AGATCGTGGAGAAAGGGTTTAGGGTTTAAAGCTCAGACTTCTATTGGAGTAAATGGGACG 420

Db 361 AGATCGTGGAGAAAGGGTTTAGGGTTTAAAGCTCAGACTTCTATTGGAGTAAATGGGACG 420

Qy 421 GTGTACATTTTCCGTTTGGAAAATGAACCTTTGGGCTTAGTTTAAATGAAGCCCATGT 480

Db 421 GTGTACATTTTCCGTTTGGAAAATGAACCTTTGGGCTTAGTTTAAATGAAGCCCATGT 480

Qy 481 TGATGGGCTTCTAGTAAATACAAATATAGTTATTTGGGCTTAGTTTAAATGAAGCCCATGT 540

Db 481 TGATGGGCTTCTAGTAAATACAAATATAGTTATTTGGGCTTAGTTTAAATGAAGCCCATGT 540

Qy 541 TGGAAATATTTGACACATGCTTGGCTACTAGTCTAAACATGCAACCGAACAGTTGTGCG 600

Db 541 TGGAAATATTTGACACATGCTTGGCTACTAGTCTAAACATGCAACCGAACAGTTGTGCG 600

Qy 601 AGACAAGTCGAGCATATACAAATGGAATGGAATGGAATGGAATGGAATGGAATGGAATGGA 660

Db 601 AGACAAGTCGAGCATATACAAATGGAATGGAATGGAATGGAATGGAATGGAATGGAATGGA 660

Qy 661 GTGTACCTTGTCTGCTGCTTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAAT 720

Db 661 GTGTACCTTGTCTGCTGCTTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAAT 720

Qy 721 ATACAAATTTTGGCTGATCTTGCACAACTCTCGATCATATCGCAATATACGTGAACA 780

Db 721 ATACAAATTTTGGCTGATCTTGCACAACTCTCGATCATATCGCAATATACGTGAACA 780

QY	781	CTGGTGATCTAAATTTGTTGTGTTAAATTTAGATTCTATCTCCGGTTTAAAG	840
Db	781	CTGGTGATCTAAATTTGTTGTGTTAAATTTAGATTCTATCTCCGGTTTAAAG	840
QY	841	TCAATATATGTTATCATCGTTTAAACATTTGTAAGTAAGATGATATAATAAATGATAAATTT	900
Db	841	TGAAATATATGTTATCATCGTTTAAACATTTGTAAGTAAGATGATATAATAAATGATAAATTT	900
QY	901	AGTTGATGGATAACGTGAAGCAAAAAATGAGATAGATACATTTTGATTTTTCGTAATTTTG	960
Db	901	AGTTGATGGATAACGTGAAGCAAAAAATGAGATAGATACATTTTGATTTTTCGTAATTTTG	960
QY	961	ACATATGGGAGAGTGAGCTACGGCATGAAGATCAAGAGACACTTGTCTCGAGCTCACAG	1020
Db	961	ACATATGGGAGAGTGAGCTACGGCATGAAGATCAAGAGACACTTGTCTCGAGCTCACAG	1020
QY	1021	AGTGACGTGTAAAAAGCTTTAGACTGAAGTCCCCATGCARACCTTAATCCTTAGTCCCATCT	1080
Db	1021	AGTGAAGTGTAAAAAGCTTTAGACTGAAGTCCCCATGCARACCTTAATCCTTAGTCCCATCT	1080
QY	1081	ACGAGAGCTCACTGACAATATATAAACTCCTCTAAGTCCCGTTCTCTTCATCCATCT	1140
Db	1081	ACCAGAGCTCACTGACAATATATAAACTCCTCTAAGTCCCGTTCTCTTCATCCATCT	1140
QY	1141	CTCACACACACAAAAAG	1158
Db	1141	CTCACACACACAAAAAG	1158
RESULT 3			
AX510242		1223 bp	DNA
LOCUS		Sequence 4937 from Patent WO0216655.	linear
DEFINITION			PAT 27-SEP-2002
ACCESSION			
VERSION		AX510242.1	GI:23391479
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
FEATURES			
source			
ORIGIN			
Query Match		100.0%;	Score 1158; DB 6; Length 1223;
Best Local Similarity		100.0%;	Pred. No. 4.3e-228;
Matches 1158; Conservative		0; Mismatches	0; Indels
			Gaps
QY	1	CACAAACATACACTCAAAATCCAGACTCACATCTACTCAATATATGCAACTTCATCATCAA	60
Db	63	CACAAACATACACTCAAAATCCAGACTCACATCTACTCAATATATGCAACTTCATCATCAA	122
QY	61	AACATCAAAACAGTCAAAAGTACAAATCAAGTCAAGTTCAGACACAAAGCCAGTAAA	120
Db	123	AACATCAAAACAGTCAAAAGTACAAATCAAGTCAAGTTCAGACACAAAGCCAGTAAA	182
QY	121	GATAGAAAAATTTAACGAAACGCTCATGCTAAGCTGCGCAAAATACCTTCCTTAATCAAAACAG	180
Db	183	GATAGAAAAATTTAACGAAACGCTCATGCTAAGCTGCGCAAAATACCTTCCTTAATCAAAACAG	242
QY	181	TAACACAGAGTAATTAGCAAAATCCGAGCAGAAAACTCTCACCCACCTTCGAAATTCACG	240

ACCESSION	AL021749
VERSION	AL021749.1
KEYWORDS	GI:2842474
SOURCE	Arabidopsis thaliana (thale cress)
ORGANISM	Arabidopsis thaliana Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
REFERENCE	1 Bevan, M., Rose, M., Hempel, S., Entian, K.-D., Hoheisel, J., Mewes, H.W., Mayer, K.F.X., Lemcke, K. and Schueller, C. Unpublished 2 (bases 1 to 93695) EU Arabidopsis sequencing project. Direct Submission Submitted (30-JUL-1999) MIPS, at the Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail: schulze@mips.biochem.mpg.de, mayer@mips.biochem.mpg.de Project Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK, E-mail: michael.bevan@bbrc.ac.uk Information on performance of analysis and a more detailed annotation of this entry and other sequences of chromosomes 3, 4 and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/.
FEATURES	Location/Qualifiers 1..93695 /organism="Arabidopsis thaliana" /mol_type="genomic DNA" /variety="Columbia" /db_xref="taxon:3702" /chromosome="4" 18..467 /gene="F2009.10" complement(18..467) /gene="F2009.10" complement(18..467) /gene="F2009.10" codon_start=1 product="hypothetical protein" protein_id="CAAL6872.1" db_xref="GI:2842475" db_xref="TrEMBL:O49442" translation="MTSYKFLFRVSKRTWACLFLVVYAILSSWNLSINNVYG ENHTSSGLPAIVASVLGVAGVLSMAALFIAPVAIVISVVTAFAGSKRR VVIHGRVKTEIAGVFRLVLLKEGNFVALCALLAIFVFNFSYSSSS" complement(18..467) /gene="F2009.10" number=1 2196..2675 /gene="F2009.20" complement(2196..2675) /gene="F2009.20" complement(2196..2675) /gene="F2009.20" codon_start=1 product="hypothetical protein" protein_id="CAAL6873.2" db_xref="GI:5701784" db_xref="TrEMBL:O49443" translation="MAKDEQETMSYKFLFRVSKRRTWCFLFVVYAVILLSSRNS LNSIVNYGENHOTSSGLPALIASVLLGAVGLVSMAALFIAPVAIVISVVTI AFPVSKRKXVIEGRKYTKELAGYVFKVLLKEGNFVALCAVIAFYFNFSYSSSS" complement(2196..2675) /gene="F2009.20" number=1 4143..6188 /gene="F2009.40" join(4143..4941,4987..5044,5096..6188) /gene="F2009.40" note="similarity to lectin receptor-like serine/threonine kinase lecrK1, Arabidopsis thaliana, PIR2:S68589 Contains Protein Kinases signatures and profile; pos. 328-349 & pos. 444-456"
gene	gene
gene	gene
CDS	CDS
exon	exon
gene	gene
gene	gene
CDS	CDS
exon	exon
gene	gene
CDS	CDS

flowers-Petunia hybrida, PATCHX:E205713
 Contains Aldhyde dehydrogenases active sites AA299-306
 Contains EST gb:R90477, R89959, AI992436.1, T04277,
 T43017, R30004
 /codon_start=1
 /product="NAM / CUC2-like protein"
 /protein_id="CAB81441.1"
 /db_xref="GI:7269708"
 /db_xref="GOA:O49461"
 /db_xref="tREMBL:O49461"
 /translation="MGLKDIKSLPPGFRPHSPDEELVCHYLCKNKIRAKSDHGVDDDD
 DDDVDEALKSGTDLVDLHICPEWELPDVAKLNKAKWYFFSFRDKYATGYRNRAT
 VSGWKATGKRTVDPRQLEDVLCRVFNKGRDSSLQDNNYNNNDQTRLEVND
 APDLNNQLPPLSSPPHNFHKKMKIQVCDQWELMKQPSRTTGHPHHCHHOTT
 ACWEQMMIGLSLSPSSHPDHESLNLLYYDNNNSVNI SGDHHQNYEKILLSLDMT
 SLDDKTCMGSSSDGMYDLHMECGGLSFETENILARQ"
 complement (11459..12034)
 /gene="AT4g28530"
 /number=1
 complement (12035..13177)
 /gene="AT4g28530"
 /number=1
 complement (13178..13332)
 /gene="AT4g28530"
 /number=2
 complement (13333..15120)
 /gene="AT4g28530"
 /number=2
 complement (15121..15337)
 /gene="AT4g28530"
 /number=3
 28030..30462
 /gene="AT4g28540"
 join(28030..28105,28237..28277,28532..28601,28760..28908,
 29004..29309,29338..29408,29503..29564,29655..29718,
 29804..29888,29989..30115,30338..30462)
 /gene="AT4g28540"
 /note="strong similarity to protein kinase
 ADK1-Arabidopsis thaliana
 Contains protein kinases signatures and profile
 AA15-38; Protein kinases signatures and profile
 contains EST gb:F19756"
 /codon_start=1
 /product="protein kinase ADK1-like protein"
 /protein_id="CAB81442.1"
 /db_xref="GI:7269709"
 /db_xref="GOA:O65463"
 Query Match 100.0%; Score 1158; DB 8; Length 197655;
 Best Local Similarity 100.0%; Pred. No. 2.4e-228;
 Matches 1158; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 CACAAACATACACTCAAAATCCAGATCACATCTACTCAATATGGAATTCATCATGAA 60
 DB 7169 CACAAACATACACTCAAAATCCAGATCACATCTACTCAATATGGAATTCATCATGAA 7288
 QY 61 AACATCAAAAACAGTCAAAAGTAAACAAATCAAGTCAGATTTCAGCACAAAGCCAGTAA 120
 DB 7229 AACATCAAAAACAGTCAAAAGTAAACAAATCAAGTCAGATTTCAGCACAAAGCCAGTAA 7288
 QY 121 GATAGAAATTTAAAGAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAA 180
 DB 7289 GATAGAAATTTAAAGAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAA 7348
 QY 181 TAAACACAGTAAATAGCAAAATCCAGCAGAAATCTCACCACCTCCGAAATTCACG 240
 DB 7349 TAAACACAGTAAATAGCAAAATCCAGCAGAAATCTCACCACCTCCGAAATTCACG 7408
 QY 241 TCTTCACTAAAAATTTTCGAAAGGAAATTCAGTCAATACCAACCCATTACACAAATACAT 300
 DB 7409 TCTTCACTAAAAATTTTCGAAAGGAAATTCAGTCAATACCAACCCATTACACAAATACAT 7468
 QY 301 TCAAAATGCGAGAATCGTACCTGGAAATCTTTCCTTCAAGTCCGAGAGAGAGAGAGAGAG 360

Db 7469 TCAAAATGCGAGAAATCGTACCTGGAAATCTTTCCTTCAAGTCCGAGAGAGAGAGAGAG 7528
 QY 361 AGATCGTGGAGAAAGGGGTTTAGGGTTTAAAGCTCAGACTTCTATTGGAGTAAATGGAGC 420
 Db 7529 AGATCGTGGAGAAAGGGGTTTAGGGTTTAAAGCTCAGACTTCTATTGGAGTAAATGGAGC 7588
 QY 421 GTGTCACATTTTCCGTTTGGAAATGAACCTTTGGGCTCACGTTATGGGCTATTAGATATT 480
 Db 7589 GTGTCACATTTTCCGTTTGGAAATGAACCTTTGGGCTCACGTTATGGGCTATTAGATATT 7648
 QY 481 TGATGGGCTTCTAGTAAATCAATATAGTTATGGCTTATGGCTTATTAATAGCCATGT 540
 Db 7649 TGATGGGCTTCTAGTAAATCAATATAGTTATGGCTTATGGCTTATTAATAGCCATGT 7708
 QY 541 TGGAAATATTGACACATGCTCTGGCTACTAGTCTTAAACATGCAACCGACAGTGTCTG 600
 Db 7709 TGGAAATATTGACACATGCTCTGGCTACTAGTCTTAAACATGCAACCGACAGTGTCTG 7768
 QY 601 AGACAAGTCGAGCATATACAAATGGATCAACACGCTTAGTGTGCGCGCTCTCGCTCAT 660
 Db 7769 AGACAAGTCGAGCATATACAAATGGATCAACACGCTTAGTGTGCGCGCTCTCGCTCAT 7828
 QY 661 GTGTCACCTGTTTCCCTCGTTTCTTAAATTTTCAAGTCTTCTTGTATTTTCA 720
 Db 7829 GTGTCACCTGTTTCCCTCGTTTCTTAAATTTTCAAGTCTTCTTGTATTTTCA 7888
 QY 721 ATACAAATTTTGGCTGTATCTTGCACAACTCTCGATCATATCGCCATATACGTGAACA 780
 Db 7889 ATACAAATTTTGGCTGTATCTTGCACAACTCTCGATCATATCGCCATATACGTGAACA 7948
 QY 781 CTGTCATCTAAATTTGTTGTTTAAATTTAGATTTCTTCTCGGTTTAAAG 840
 Db 7949 CTGTCATCTAAATTTGTTGTTTAAATTTAGATTTCTTCTCGGTTTAAAG 8008
 QY 841 TGAATTTATGTTATCATGTTTAAACATTTAGTAAAGTAAATGATAAATAAATTT 900
 Db 8009 TGAATTTATGTTATCATGTTTAAACATTTAGTAAAGTAAATGATAAATAAATTT 8068
 QY 901 AGTTGATGTAACGTTGAACAAAATAGATAGATACATTTGATTTTCTCGTATTTTG 960
 Db 8069 AGTTGATGTAACGTTGAACAAAATAGATAGATACATTTGATTTTCTCGTATTTTG 8128
 QY 961 ACATATGCGAGAGTGAAGTACCGCATGAAGTCAAGACACTTGTCTCGACTCACAG 1020
 Db 8129 ACATATGCGAGAGTGAAGTACCGCATGAAGTCAAGACACTTGTCTCGACTCACAG 8188
 QY 1021 AGTGACGTGTAAAAGCTTAGACTGAAGTCCCATGCAACCTTAATCTTACGTGGTCAA 1080
 Db 8189 AGTGACGTGTAAAAGCTTAGACTGAAGTCCCATGCAACCTTAATCTTACGTGGTCAA 8248
 QY 1081 ACCACGAGCTCACTTGACAATATATAAATCTCTTCAAGTCCCGTCTCTTCTCATCTCT 1140
 Db 8249 ACCACGAGCTCACTTGACAATATATAAATCTCTTCAAGTCCCGTCTCTTCTCATCTCT 1140
 QY 1141 CTCACAACAAACAAAAG 1158
 Db 8309 CTCACAACAAACAAAAG 8326
 RESULT 6
 ATU66594/c
 LOCUS
 DEFINITION Arabidopsis thaliana 3054 bp DNA linear PLN 05-JAN-1999
 ACCESSION U66594
 VERSION U66594.1
 KEYWORDS Arabidopsis thaliana (thale cress)
 SOURCE Arabidopsis thaliana
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 REFERENCE 1 (bases 1 to 3054)

889	TGATGGGCTTCTAGTAAATACAAATATAGTATTTGGGCTTTGGTATTTAAATGAAAGCGGCTG	
541	TGGAATATTTGACACATGCTCTTTGGCTACTAGTGTCTAAACATGCCAACCGAAACAGTTGTGCG	600
829	TGGAAATATTTGACACATGCTCTTTGGCTACTAGTGTCTAAACATGCCAACCGAAACAGTTGTGCG	770
601	AGACAAGTCGCAGCATATACAATGGATCAACACGCCCTAGTGTGCGCGGCTCTCGCTCAT	660
769	AGACAAGTCGCAGCATATACAATGGATCAACACGCCCTAGTGTGCGCGGCTCTCGCTCAT	710
661	GTGTCACTGTTTTCCTCG-TTTTTTTTAAATTTTTCATAAGTCTCTTTTGTGTTTATCTTC	719
709	GTGTCACTGTTTTCCTCGTTTTTTTTTAAATTTTTCATAAGTCTCTTTTGTGTTTATCTTC	650
720	AATACAAATTTTTTGGCTGTATCTTGCAAACTCTTCGATCATATCGCCAAATATACGTGAAC	779
649	AATACAAATTTTTTGGCTGTATCTTGCAAACTCTTCGATCATATCGCCAAATATACGTGAAC	590
780	ACTGGTGATCTAAATTTGTTGTTTAATCTGTTAAATTTAGATTTCTATCTCCGGTTTAAAA	839
589	ACTGGTGATCTAAATTTGTTGTTTAATCTGTTAAATTTAGATTTCTATCTCCGGTTTAAAA	530
840	GTGAATTTATATGATCATGCTTAAAAACATTTGAAGTAAGATGATAATAAAATGATAAAT	899
529	GTGAATTTATATGATCATGCTTAAAAACATTTGAAGTAAGATGATAATAAAATGATAAAT	470
900	TAGTTGATGGATTAACGTGAAGCAAAAAATTGAGATAGATACATTTGTTGTCGTATTTT	959
469	TAGTTGATGGATTAACGTGAAGCAAAAAATTGAGATAGATACATTTGTTGTCGTATTTT	410
960	GACATATCCGGAGAGTCAGCTACGCGCATGAAGATCAAGAGACACTTGTCTCGAGCTCACA	1019
409	GACATATCCGGAGAGTCAGCTACGCGCATGAAGATCAAGAGACACTTGTCTCGAGCTCACA	350
1020	GAGTGACGTGTAAAAAGCTTTAGACTGAAGTCCCACTGCCAAAACCTAACTCAGTGGCTCA	1079
349	GAGTGACGTGTAAAAAGCTTTAGACTGAAGTCCCACTGCCAAAACCTAACTCAGTGGCTCA	290
1080	AACGACGAGCTCACTTGACAAATATATAACTCTCTTAAGTCCCGTTCTTCTCATCCATC	1139
289	AACGACGAGCTCACTTGACAAATATATAACTCTCTTAAAGTCCCGTTCTTCTCATCCATC	230
1140	TCTCACAAACAAAAAG 1158	
229	TCTCACAAACAAAAAG 211	
RESULT 7		
ATU66916		
LOCUS	3100 bp DNA linear	PLN 23-OCT-1999
DEFINITION	Arabidopsis thaliana 12S cruciferin seed storage protein (ATCRU3)	
ACCESSION	U66916	
VERSION	U66916.1	
KEYWORDS	GI:1628582	
SOURCE	Arabidopsis thaliana (thale cress)	
ORGANISM	Arabidopsis thaliana	
REFERENCE	1 (bases 1 to 3100)	
AUTHORS	McCourt,P., Ferraioli,G. and Riggs,C.D.	
TITLE	Characterization of a cruciferin-deficient mutant of Arabidopsis	
JOURNAL	Unpublished	
REFERENCE	2 (bases 1 to 3100)	
AUTHORS	Riggs,C.D.	
TITLE	Direct Submission	
JOURNAL	Submitted (15-AUG-1996) Botany, University of Toronto, 1265 Military Trail, Scarborough, Ontario M1C 1A4, Canada	
FEATURES	Location/Qualifiers	
source	1.. 3100	
	/organism="Arabidopsis thaliana"	
	/mol_type="genomic DNA"	


```

/db_xref="taxon:3702"
/chromosome="4"
929. .3039
/gene="ATCRU3"
TATA_signal
929. .934
/gene="ATCRU3"
CDS
join(988. .1291,1417. .1841,1947. .2384,2602. .3009)
/gene="ATCRU3"
/codon_start=1
/product="12S cruciferin seed storage protein"
/protein_id="AAB17379.1"
/db_xref="GI:1628583"
/translation="MVKLSNLLVATFGLVLLINGCLARQSLGVPQLQNECNLDLV
LQATETIKSEAGLEIYDHNHPQLRCVGSVARYVIEOGGLYLPTFFSPKISYVVOG
TGISRVVPGQAEITFMDSPQMGQOGQOPWQOGQOGQOPWQOGQOGQOGQOGQOPWE
GQOGQOGQOGQOGQOGQOGQOGQOGQOGQOGQOGQOGQOGQOGQOGQOGQOGQOGQ
GEOPVILFALLDIANYONQDLNRPRVFLHAGNQQGGGGGQOGQOGQOGQOGQOGQOG
IAQALKDVLQALQONQDGRNIVRVKGFVVRPELRQFYES EEMRHPRSQONG
LEETICSRSHENIDDPADRVKPSLGRVTSVNSYTLPILEYVRLSATRGVLQGNM
VLPKYNMNAEILYCTGGQRIQVNDGQNVLDQVQKGLVVIPOGFVAVVOSHGN
KFEWISPTNENAMISTLAGRTSLRLPLEVINSNGFQISPEEARKIKFNTLETILR
AAGRQOOLIEIVEA"
<988. .1291
/gene="ATCRU3"
exon
number=1
1417. .1841
/gene="ATCRU3"
exon
number=2
1947. .2384
/gene="ATCRU3"
exon
number=3
2602. .3039
/gene="ATCRU3"
exon
number=4
3034. .3039
polya_signal
/gene="ATCRU3"
ORIGIN

```

Query Match 82.6%; Score 956.2; DB 8; Length 3100;
Best Local Similarity 99.5%; Pred. No. 1.3e-196;
Matches 969; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

```

186 ACAGTAATTAGCAAAATCCGAGAGAAACTCTCACCCACCTCCGAAATTCACGCTTC 245
11 ACAGTAATTANCTAAATCCGAGAGAAACTCTCACCCACCTCCGAAATTCACGCTTC 70
246 ACTAAATTTTCCGAAGGATCGATCAATACCAACCACTTACACAAATACATAATCAAA 305
71 ACTAAATTTTCCGAAGGATCGATCAATACCAACCACTTACACAAATACATAATCAAA 130
306 ATGGCAGAAATCGTACCTGGAAACTTTGCTTCAAGTCGAGAGAGAGGAAAGGAGATC 365
131 ATGGCAGAAATCGTACCTGGAAACTTTGCTTCAAGTCGAGAGAGAGGAAAGGAGATC 190
366 GTGGAGAAAGGGTTTAGGGTTTAAGCTCAGACTTCTATTGGAGTAAATGGACGGTGTC 425
191 GTGGAGAAAGGGTTTAGGGTTTAAGCTCAGACTTCTATTGGAGTAAATGGACGGTGTC 250
426 ACATTTTCGGTTTGGAAATGAACCTTGGCTCAGCTTATGGCTATTAAGATTTTGGATG 485
251 ACATTTTCGGTTTGGAAATGAACCTTGGCTCAGCTTATGGCTATTAAGATTTTGGATG 310
486 GGCTTTCTAGTAAATACATATAAGTTATTGGCTTAGTTTAAATTAAGCCCATGTTGGAA 545
311 GGCTTTCTAGTAAATACATATAAGTTATTGGCTTAGTTTAAATTAAGCCCATGTTGGAA 370
546 ATATTGACACATGTTTGGCTTACTAGTCTTAAACATGCAACCGAACACTTGTGAGACA 605
371 ATATTGACACATGTTTGGCTTACTAGTCTTAAACATGCAACCGAACACTTGTGAGACA 430
606 AGTCGACGATATACATGATCAACACGCTAGTGTGCGCGCTCGCTCATGTGTGTC 665
431 AGTCGACGATATACATGATCAACACGCTAGTGTGCGCGCTCGCTCATGTGTGTC 490

```

```

QY 666 ACCTTGTTCCTCG-TTTTTTTTAAATTTTTCATAAGTCTCTTTTGTATCTTCAATAC 724
Db |||||
QY 491 ACCTTGTTCCTCGTTTTTTTTTTTAAATTTTTCATAAGTCTCTTTTGTATCTTCAATAC 550
Db |||||
QY 725 AAATTTTGGCTGTATCTTGCNAACCTCTCGATCATATCGCAATATACGTGAAACACTGG 784
Db |||||
QY 551 AAATTTTGGCTGTATCTTGCNAACCTCTCGATCATATCGCAATATACGTGAAACACTGG 610
Db |||||
QY 785 TGATCTAAATTTGTGTGTTAAATTTTAAATTTTAGATTCTATCTCCGTTTAAAGTGAA 844
Db |||||
QY 611 TGAICTAATTTGTGTGTTAAATTTTAAATTTTAGATTCTATCTCCGTTTAAAGTGAA 670
Db |||||
QY 845 TTATGATCATCGTTTAAATTTTAAATTTTAAATTTTAGATTCTATCTCCGTTTAAAGTGAA 704
Db |||||
QY 671 TTATGATCATCGTTTAAATTTTAAATTTTAAATTTTAGATTCTATCTCCGTTTAAAGTGAA 730
Db |||||
QY 905 GATGATAACCTGAGCAAAAATTTAGATGATATATTTGATTTTGTCTGTTTGTGACAT 964
Db |||||
QY 731 GATGATAACCTGAGCAAAAATTTAGATGATATATTTGATTTTGTCTGTTTGTGACAT 790
Db |||||
QY 965 ATGCGGAGAGTACGCTACGCGCATGAAGATCAAGAGACACTTGTCTCGAGCTCAGAGTG 1024
Db |||||
QY 791 ATGCGGAGAGTACGCTACGCGCATGAAGATCAAGAGACACTTGTCTCGAGCTCAGAGTG 850
Db |||||
QY 1025 ACCTGTTAAAAGCTTAGACTGAACTCCCATGCAAACTTAATCTACGTGCTCAAAACA 1084
Db |||||
QY 851 ACCTGTTAAAAGCTTAGACTGAACTCCCATGCAAACTTAATCTACGTGCTCAAAACA 910
Db |||||
QY 1085 CGAGCTCAGTGAATATATATAATCTCTCTTAAGTCCCGTTCTCTTCAATCCATCTCTCA 1144
Db |||||
QY 911 CGAGCTCAGTGAATATATATAATCTCTCTTAAGTCCCGTTCTCTTCAATCCATCTCTCA 970
Db |||||
QY 1145 CAACAACAAAAG 1158
Db |||||
QY 971 CAACAACAAAAG 984

```

```

RESULT 8
BX296428/c
LOCUS
DEFINITION
  Arabidopsis thaliana transposon insertion STS SM_3.33744, sequence
  tagged site.
ACCESSION
  BX296428
VERSION
  STS; STS, sequence tagged site.
KEYWORDS
  Arabidopsis thaliana (thale cress)
SOURCE
  Arabidopsis thaliana
  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
  Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
  rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE
  1
  Clarke J.H., Bowles B., Carter J., Hart D., McCullagh B.,
  Murphy G., Langham S., LeGrys C., Jones J.D.G. and Bevan M.
  Unpublished
  2 (bases 1 to 329)
  Clarke J.H.
  Direct Submission
  Submitted (24-MAR-2003) Clarke J.H., John Innes Centre, Colney
  Lane, Norwich, NR4 7UJ, UK
  AT denotes an activation tag dissociation transposon within a
  single line, BT an enhancer trap dissociation transposon, GT a gene
  trap dissociation transposon, MT a mis-expression enhancer trap
  dissociation transposon, SM a defective suppressor mutator
  transposon. 3 denotes a sequence derived from the 3' end of the
  transposon, 5 denotes a sequence derived from the 5' end of the
  transposon. BSRG GARNET, ATIS project
  On-line seed stock requests: http://nasc.nott.ac.uk/ NASC stock
  code: DEAD.
  Location/Qualifiers
    1..329
    /organism="Arabidopsis thaliana"
    /mol_type="genomic DNA"
FEATURES
  source

```


QY 960 GACATATCGGAGAGTACGCTACGCGCATGAAGATCAAGAGACACTTCTCGAGCTCACA 1019
 Db |||||
 1046 GGAGTATCGGAGATAGTCGACGCGCATGAACATCATGAGACACTTCTCGAGCTCACA 1105
 QY 1020 GAGTACCTGTAAAGCTTACGATGAGTCCCAATGCAAACTTAATCTAGTGGCTCA 1079
 Db |||||
 1106 GAGTGACGTGTAAGACCATGACCCAGCACTTATGCAAACTTCTAGTGGCACA 1165
 QY 1080 AACCACGAGCTCACTTGACATATATAAATCTCTTAAG--TCCGTTCTCTTCATCCA 1137
 Db |||||
 1166 AACCTTATGCTCACTCCACATATATAAATCTCTACCAAGTCTCCATGTTCTTTCATCCA 1225
 QY 1138 TCTCTCAACAAACAAAA 1157
 Db TCTATCAACAAACACACAA 1245

RESULT 10
 AF319771
 LOCUS
 DEFINITION Brassica napus CUC2-like protein gene, partial cds; and prohibitin 1-like protein and cruciferin subunit genes, complete cds.
 ACCESSION AF319771
 VERSION AF319771.1 GI:12751301
 KEYWORDS
 SOURCE
 ORGANISM Brassica napus (rape)
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
 1 (bases 1 to 6049)
 Scherthanner, J.P. and Robert, L.S.
 Genomic clone from Brassica napus containing genes encoding a cruciferin subunit, a prohibitin 1-like protein and the carboxy terminus of a CUC2-like protein
 Unpublished
 2 (bases 1 to 6049)
 Scherthanner, J.P. and Robert, L.S.
 Direct Submission
 Submitted (06-NOV-2000) Eastern Cereal and Oilseed Research Center, Agriculture and Agri-Food Canada, 960 Carling Ave. KWN 2022, Ottawa, Ontario K1A 0G6, Canada
 Location/Qualifiers
 1..6049
 /organism="Brassica napus"
 /mol_type="genomic DNA"
 /cultivar="Westar"
 /db_xref="taxon:3708"
 <54..>625
 /product="CUC2-like protein"
 <54..625
 /codon_start=3
 /product="CUC2-like protein"
 /protein_id="AAK07611.1"
 /db_xref="GI:12751304"
 /translation="VNDGTVRHFHPSKAYTTGTRKRVKHTDSEVDGETRWKGTG KTRPVLTGSRVGYKILVLYNYGKQKPKENTMWRQYHLGTNEEKEBELVSVKY FYQTQPCGGGSAVAATKPERPYLHGGSHLGGGGRHLHYLHNVKSGGSAVAGE YNNIPALISFNQGTQNHVNGSQFFIP"
 complement(join<1106..1300,1388..1600,1686..1856, 1998..2114,2377..>2553))
 /product="prohibitin 1-like protein"
 complement(join(1106..1300,1388..1600,1686..1856, 1998..2114,2377..2553))
 /codon_start=1
 /product="prohibitin 1-like protein"
 /protein_id="AAK07610.1"
 /db_xref="GI:12751303"
 /translation="WNLNKVPKVPGGGAVSALLKIGIIGSLGYGATHSLYNDVGG HRAINFRLVGKQVPEGTGLMVFEPFVYDVRARPLYVESTSGSRDIQMWKIG LRLVLRPMADQLPEYRTIGENYSRVLPSIIHETLKAVQAQNASOLITOREAVSRE TRKILTORATNFNIADDVSIITLTFGKEFTAAIEAKVAAQAEARAKFIVEKABQDK

FEATURES
 source

mRNA

CDS

mRNA

CDS

misc_structure
 TATA_signal
 mRNA
 CDS

polyA_signal
 ORIGIN

Query Match 16.7%; Score 193; DB 8; Length 6049;
 Best Local Similarity 68.3%; Pred. No. 1.1e-29;
 Matches 408; Conservative 0; Mismatches 160; Indels 29; Gaps 9;
 QY 568 ACTAGTCTAAACATGCAACCGACAGTTCGAGACAACTCGCAGATATACAAATGGAT 627
 Db |||||
 3399 ACTGCTCTTAAATGCAACCGACAGTTCGAGACAACTCGCAGAGAT-CAACGGAC 3457
 QY 628 CAAACAGCCTAGTGTGCGCGCTCTC-----GCTCATGTGCACCTTGTCTCTCTTT 682
 Db |||||
 3458 CAAACCTCTAGTGTGCGCGCTCTCNAAGTCTCAAGTGTCACTTTTGTCTCTTT 3517
 QY 683 TTTTAAATTTTCAAGTCTCTTTGTTTATCTTCAATAACAAATTTTGGCTGTATCT 742
 Db |||||
 3518 CCTTTGAATTT-----GGTAACTCTTTCAATAATATATATAGTTTGGCTGTATCT 3566
 QY 743 TGCAAACTCTTCGATCATATCGCCA-----ATATACGTGAACACTGGTGAATTAATTTG 796
 Db |||||
 3567 CCTAGACTGTTCGATCACAACCCAGATATTTAATACTGGATACCTAGTGAATTTG 3626
 QY 797 TTGTGTTAATTTG-TAAATTTAGATTCTATCTCCGTTTAAAGTGAATTAATGATC 855
 Db |||||
 3627 ATAGCTTAATTTGCAAAAAGTAGATTCTCTATCCCTTTATTATATGTTAAATGAAT 3686
 QY 856 ATGGTTAAACATTTGTAAGTGAATGAATAATAAATGATAAAATTTAGTTGATGATAACG 915
 Db |||||
 3687 TAAGACGATTAAAAATGAAATTAATCA-CTAGAGAGATTCCTCTGCTGTTGATACCA 3745
 QY 916 TGAAGCAAAAATGAGATAGATACATTTGATTTTGTCTGATTTTGCATATGCGGAGAT 975
 Db |||||
 3746 TGAACCTAAAATGAGAGATACA-TTTGAAATTTTGTCTGTTGAGATATCGGAGATA 3804
 QY 976 GAGCTACGCGCATGAAGATCAAGAGACACTCTCGATGCTCAGAGTCAAGTGAATGAAAAA 1035
 Db |||||
 3805 GTGCTACGCGCATGAACATCATGAGACACTTGCCTTACAGTCAAGTGAATGCGGAGATA 3864
 QY 1036 GCTTACGTAAGTCCCATGCAAACTTAATCTTACGTGGCTCAAAACCAAGCTCACTT 1095
 Db |||||
 3865 CCATAGACCCAGCTTTCATGCAAACTAGTTCTTACGTGGCAAACTTTCATGCTCACTC 3924
 QY 1096 GACATATATAAATCTCTCTTAAG--TCCGTTCTCTTCATCCATCTCTCAACAA 1150
 Db |||||
 3925 CAC-ATATATAAATCTCTACCAAGTCTCCATGTTTCTTTCATCCATCTATCACAAA 3980

RESULT 11

RSCRUG

LOCUS

DEFINITION Raphanus sativus pgCruR5E5 gene for cruciferin precursor.
 4901 bp DNA linear PLN 06-NOV-2003

ACCESSION X59808 S47559
VERSION X59808.1 GI:21117
KEYWORDS 12S globulin; class II; cruciferin; seed storage protein.
SOURCE Raphanus sativus (radish)
ORGANISM Raphanus sativus
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Raphanus.
REFERENCE 1 (bases 1 to 4901)
Depigny-Thijs, D., Raynal, M., Aspart, L., Delseny, M. and Grellet, F.
The cruciferin gene family in radish
Plant Mol. Biol. 20 (3), 467-479 (1992)
MEDLINE 93043037
PUBMED 1421150
REFERENCE 2 (bases 1 to 4901)
Depigny D.M.C.
Direct Submission
Submitted (26-MAY-1991) Depigny D.M.C., University of Perpignan,
CNRS UA 565, Lab. de Physiologie Vegetale, Ave de Villeneuve, 66860
Perpignan, Cedex, FRANCE
COMMENT On Nov 19, 2003 this sequence version replaced gi:259697.
FEATURES
source
1..4901
/organism="Raphanus sativus"
/mol_type="genomic DNA"
/cultivar="Saxa Knacker"
/db_xref="taxon:3726"
/clone="CruR4"
TATA_signal
678..685
704..2676
/gene="pgCruR4"
join(704..1043,1187..1470,1568..2014,2272..2676)
mRNA
/gene="pgCruR4"
/product="cruciferin precursor"
/evidence="experimental"
704..1043
/gene="pgCruR4"
/number=1
join(740..1043,1187..1470,1568..2014,2272..2676)
CDS
/gene="pgCruR4"
/codon_start=1
/product="cruciferin precursor"
/protein_id="CAA42478.1"
/db_xref="GI:21117"
/db_xref="GOA:Q02498"
/translation="MVLAHLVATFGVLLVGLGRLQSLGVPVQLGNACNLNDLV
LQPTIKSEAGRLVYDHPHQLRCAGVSRLIIEOGGLVLPFFSSPKIAYVVG
TAGSAHWIYNTGDDPLVIVCLLDIANYQNLDNRPTFLRAGNPGGSHQQQQQQQ
NMLSGFDQVLAQLKMLRLAQLQNDNRNIVRYKFPVVRPLRQYSEQW
RHPRGPDSODNGLEETICSMTHENIDDPADRVYKPNLGRVTSVNSYLPILQYI
PLSATRGLIQNMVLPKYNMNEILYCTQOARIQVNDNGQNLVDQVQKGLV
IPQFAYVYQSHGNFNWISPTNANMVSTLGRTSALRALPLEVITNAFQISLEEA
RRIPFNTPTLITLTHARGQPLIEVEA"
740..808
/gene="pgCruR4"
join(809..1043,1187..1470,1568..1846)
mat_peptide
/gene="pgCruR4"
/product="cruciferin alpha subunit"
join(1847..2014,2272..2673)
mat_peptide
/gene="pgCruR4"
/product="cruciferin beta subunit"
1044..1186
/gene="pgCruR4"
/number=1
join(1187..1470
/gene="pgCruR4"
/number=2
1471..1567
/gene="pgCruR4"
/number=2
1568..2014
exon

intron
/gene="pgCruR4"
/number=3
2015..2271
/gene="pgCruR4"
/number=3
2272..2676
/gene="pgCruR4"
/number=4
2699..2704
polyA_signal
2829..2834
polyA_signal
3002..3007
ORIGIN
Query Match 15.6%; Score 181; DB 8; Length 4901;
Best Local Similarity 63.7%; Pred. No. 3.5e-27;
Matches 392; Conservative 0; Mismatches 180; Indels 43; Gaps 6;
QY 574 GCTAAACATGCAACCGAACAGTTGTGCGAGACAGTGTGCGAGATCGCAGCATATATCAATGGATCAACA 633
DB 135 GCTAAACATGCAACCGAACAGTTGTGCGAGATCGCAGATCAAGTCGAGAAGATAAAACGACCAACCC 194
QY 634 CGCTAGTGTGCGGGGTCTCGCTCATGTGTCACTTGTTCCTCGTTTCTCTCTTTTCTTTTAAATTT 693
DB 195 CTCGGATGTGCGGGGTCTCA--TCAAGTGTCACTTTTGTCTCTTTTCTTTTCTTTTGAATTT 252
QY 694 TCCATAAGTTCTTTTGTGTTTATCTTCAATACAAATTTTGGCTGTATCTTGCACAACTTT 753
DB 253 GGTAACTTTTGTGTTTATCTTCAATACAAATTTTGGCTGTATCTTGCACAACTTT 302
QY 754 CGATCATATCGCCA-----ATATACGTGAACACTGTGTGATCTTAATTTGTGTTTAAATTT 807
DB 303 CGATTGCGAGCTAGATATTTTAAATCTCGACACTAGTGTGATGATTTGATTTGTTTAAATTT 362
QY 808 GTTAAATTTAGATTTCTATCTCCGGTTTAAAGTGAATTTATGTATCATGTGTTTAAACA 867
DB 363 GTCAAAAGTAGATTTCCCATATGACCATATCTCGTTTACTCACTACTATATGATTAA 422
QY 868 TTGTAAGTAAAGTAAATATAA-----ATGATAAATTTAGTTGA 906
DB 423 ATGAATTTAGACGACTACAAATGATTAATCAATCAACAGAGTTGCTTACTATGATAA 482
QY 907 TGGATAACGTGAACCAAAATAGATAGATATGATTTG-ATTTGTGATTTTGTGACATA 965
DB 483 GTGATAACATGAACCTTAAATAGATAGATATGATTTGAAATTTTGTGTTTGGAGTA 542
QY 966 TCGGAGAGTGTGCTAGCGCATGAGATCAAGAGACTTGTCTGAGCTCACAGATGA 1025
DB 543 TCGGAGATTTGTGCTAGCGCATCAACATTTGTGAGACACTTGTGTTGAACTCACAGATGA 602
QY 1026 CGTGTAAAGCTTAGACTGAAGTCCCATGCAACCTTAATCTCTACGTGGCTCAACCCAC 1085
DB 603 CGTGTAAACCATAGACCCAGACTTCAATGCAACCCACTCTACGTGGCAAAACCTT 662
QY 1086 GAGCTCACTTGACAATATATAAATCTCTCTCTTAAGT---CCGTTCTCTTCACTCTCT 1142
DB 663 CATGCTCACTCCACATATATAACCCCTACTAAGTCCACCTACTTCTTCTTCTCTCT 722
QY 1143 CACAACAACAAA 1157
DB 723 CACAACAACAAA 737

RESULT 12
BX545778/c
LOCUS BX545778.1
DEFINITION Arabidopsis thaliana transposon insertion STS SM_3.36286, sequence tagged site.
ACCESSION BX545778
VERSION BX545778.1 GI:32169011
KEYWORDS STS; STS, sequence tagged site.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

REFERENCE	AUTHORS	JOURNAL	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1	Clarke, J.H., Bowles, B., Carter, J.J., Hart, D., McCullagh, B., Walsh, S., Langham, S., LeGrys, C., Jones, J.D.G. and Bevan, M.	Unpublished	2	(bases 1 to 150)	Clarke, J.H.	Direct Submission	Submitted (23-JUN-2003) Clarke J.H., John Innes Centre, Colney Lane, Norwich, NR4 7UJ, UK
AT	denotes an activation tag dissociation transposon within a single line, ET an enhancer trap dissociation transposon, GT a gene trap dissociation transposon, MT a mis-expression enhancer trap dissociation transposon, SM a defective suppressor mutator transposon. 3 denotes a sequence derived from the 3' end of the transposon. 5 denotes a sequence derived from the 5' end of the transposon. BSRG GARNET, ATIS project	On-line seed stock requests: http://nasc.nott.ac.uk/ NASC stock code: N122997.					
FEATURES	source						
ORIGIN							
Query Match	13.0%;	Score 150;	DB 11;	Length 150;			
Best Local Similarity	100.0%;	Pred. No. 1.2e-20;					
Matches 150;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;			
Qy	895	AAATTTAGTTGATGATGAACGTGAGCAAAAATGAGATAGATACATTTGTTGCTG	954				
Db	150	AAATTTAGTTGATGATGAATGCTGAGCAAAAATGAGATAGATACATTTGTTGCTG	91				
Qy	955	ATTTTGACATATGCGGAGAGTGAGCTACGCGCATGAGATCAAGACACATTTGCTCGAGC	1014				
Db	90	ATTTTGACATATGCGGAGAGTGAGCTACGCGCATGAGATCAAGACACATTTGCTCGAGC	31				
Qy	1015	TCACAGAGTGACGTGTAAAAAGCTTAGACT	1044				
Db	30	TCACAGAGTGACGTGTAAAAAGCTTAGACT	1				
RESULT 13	BX296260/c						
LOCUS	180 bp	DNA	linear	STS 10-JUN-2003			
DEFINITION	Arabidopsis thaliana transposon insertion	STS SM_3.23228					
ACCESSION	BX296260						
VERSION	BX296260.1	GI:29170125					
KEYWORDS	STS; STS						
SOURCE	Arabidopsis thaliana						
ORGANISM	Arabidopsis thaliana						
REFERENCE							
AUTHORS	Clarke, J.H., Bowles, B., Carter, J.J., Hart, D., McCullagh, B., Murphy, G., Langham, S., LeGrys, C., Jones, J.D.G. and Bevan, M.						
JOURNAL	Unpublished						
REFERENCE	2	(bases 1 to 180)					
AUTHORS	Clarke, J.H.						
TITLE	Direct Submission						
JOURNAL	Submitted (24-MAR-2003) Clarke J.H., John Innes Centre, Colney Lane, Norwich, NR4 7UJ, UK						
COMMENT	AT denotes an activation tag dissociation transposon within a single line, ET an enhancer trap dissociation transposon, GT a gene trap dissociation transposon, MT a mis-expression enhancer trap dissociation transposon, SM a defective suppressor mutator transposon. 3 denotes a sequence derived from the 3' end of the transposon. 5 denotes a sequence derived from the 5' end of the transposon. BSRG GARNET, ATIS project						

```

/protein_id="AA05593.1"
/db_xref="GI:21593626"

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2004, 07:01:07 ; Search time 816 Seconds
(without alignments)
7219.177 Million cell updates/sec

Title: US-09-998-059-1
Perfect score: 1158
Sequence: 1 cacaacatacactcaaaat.....ctctcacaacacacaaaaag 1158

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 3372238 seqs, 2543544081 residues

Total number of hits satisfying chosen parameters: 6744476

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/2/pubpna/US09E_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
- 12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
- 13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
- 14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
- 15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
- 16: /cgn2_6/ptodata/2/pubpna/US10D_PUBCOMB.seq:*
- 17: /cgn2_6/ptodata/2/pubpna/US10E_PUBCOMB.seq:*
- 18: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
- 19: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
- 20: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1158	100.0	1158	10	US-09-998-059-1
2	1158	100.0	1164	10	US-09-998-059-13
3	1158	100.0	1223	9	US-09-938-842A-4937
4	1158	100.0	1233	11	US-09-938-842A-4937
5	65.4	5.6	10326	15	US-10-311-455-2139
6	64.2	5.5	547	17	US-10-311-455-16890
7	56.4	4.9	433	17	US-10-021-323-16890
8	56.2	4.9	6000	15	US-10-311-455-1106
9	56.2	4.9	40324	17	US-10-433-793-179
10	55.8	4.8	5845	15	US-10-311-455-1635
11	54.6	4.7	7441	16	US-10-257-166-140
12	54.6	4.7	7479	16	US-10-240-454-40
13	54.2	4.7	6713	15	US-10-311-455-1053

C 14	54.2	4.7	6713	15	US-10-240-485-79	Sequence 79, Appl
C 15	53.4	4.6	520	17	US-10-021-323-7699	Sequence 7699, Ap
C 16	53	4.6	3673778	15	US-10-312-841-1	Sequence 1, Appli
C 17	52.2	4.5	337	9	US-09-960-352-6976	Sequence 6976, Ap
C 18	51.2	4.4	529	17	US-10-021-323-10810	Sequence 2, Appli
C 19	51.2	4.4	96588	11	US-09-997-722-292	Sequence 292, App
C 20	51.2	4.4	3673778	15	US-10-312-841-2	Sequence 10810, A
C 21	51	4.4	11735	14	US-10-239-676-33	Sequence 2, Appli
C 22	51	4.4	11735	15	US-10-240-453-41	Sequence 33, Appli
C 23	51	4.4	11735	16	US-10-221-613-99	Sequence 41, Appli
C 24	51	4.4	13511	15	US-10-311-455-253	Sequence 99, Appli
C 25	50.6	4.4	8011	15	US-10-311-455-51	Sequence 253, App
C 26	50.4	4.4	419	9	US-09-960-352-11234	Sequence 51, Appli
C 27	50.4	4.4	7544	14	US-10-239-676-5	Sequence 11234, A
C 28	50.4	4.4	7544	15	US-10-240-453-13	Sequence 5, Appli
C 29	50.2	4.3	7037	15	US-10-311-455-1724	Sequence 13, Appli
C 30	50.2	4.3	13503	15	US-10-311-455-2043	Sequence 1724, Ap
C 31	50	4.3	560	17	US-10-021-323-2253	Sequence 2043, Ap
C 32	50	4.3	5241	16	US-10-221-613-414	Sequence 2253, Ap
C 33	49.8	4.3	6227	15	US-10-311-455-1559	Sequence 414, App
C 34	49.8	4.3	6227	16	US-10-240-589C-69	Sequence 1559, Ap
C 35	49.8	4.3	6476	15	US-10-311-455-514	Sequence 69, Appli
C 36	49.8	4.3	3673778	15	US-10-312-841-2	Sequence 514, Appli
C 37	49.6	4.3	499	17	US-10-021-323-10189	Sequence 2, Appli
C 38	49.6	4.3	6104	15	US-10-311-455-269	Sequence 10189, A
C 39	49.6	4.3	6104	16	US-10-221-714A-17	Sequence 269, App
C 40	49.6	4.3	83391	17	US-10-433-793-123	Sequence 17, Appli
C 41	49.4	4.3	7238	15	US-10-311-455-421	Sequence 123, App
C 42	49.4	4.3	17142	14	US-10-239-676-205	Sequence 421, App
C 43	49.4	4.3	17142	15	US-10-311-455-2079	Sequence 205, App
C 44	49.4	4.3	17142	15	US-10-240-453-303	Sequence 2079, Ap
C 45	49.4	4.3	17934	15	US-10-311-455-1692	Sequence 303, App

ALIGNMENTS

RESULT 1

US-09-998-059-1
; Sequence 1, Application US/0998059
; Publication No. US20030005485A1
; GENERAL INFORMATION:
; APPLICANT: Ohlrogge, John B.
; APPLICANT: Benning, Christoph
; APPLICANT: Gao, Hongbo
; APPLICANT: Girke, Thomas
; APPLICANT: White, Joseph A.
; TITLE OF INVENTION: Plant Seed Specific Promoters
; FILE REFERENCE: MSU-06689
; CURRENT APPLICATION NUMBER: US/09/998,059
; CURRENT FILING DATE: 2001-11-30
; PRIOR APPLICATION NUMBER: 60/250,401
; PRIOR FILING DATE: 2000-12-01
; NUMBER OF SEQ ID NOS: 60
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1158
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-998-059-1

Query Match
Best Local Similarity 100.0%; Score 1158; DB 10; Length 1158;
Matches 1158; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	CACAAACATACACTCAAAATCCAGACTCACATCTACTCAATTTATGCAACTTCATCATGAA	60
DB	1	CACAAACATACACTCAAAATCCAGACTCACATCTACTCAATTTATGCAACTTCATCATGAA	60
QY	61	AACATCAAAAACAGTCAAAAGTAACAATCAAGTCAGATTCAGCAACAAGCCAGTAAA	120
DB	61	AACATCAAAAACAGTCAAAAGTAACAATCAAGTCAGATTCAGCAACAAGCCAGTAAA	120

Thu Oct 14 09:37:08 2004

us-09-998-059-1.rnpb

```
QY 121 GATAGAAAAATTTAACGAAACGCTCATCTAGCTGCGCAAAAATATCTTCCTTAATCAAAACAG 180
DB 121 GATAGAAAAATTTAACGAAACGCTCATCTAGCTGCGCAAAAATATCTTCCTTAATCAAAACAG 180
QY 181 TAAACAAGAGTAAATAGCAAAATCCGAGCAGAAAGCTCTCACCCACCTCCGAAATTCACG 240
DB 181 TAAACAAGAGTAAATAGCAAAATCCGAGCAGAAAGCTCTCACCCACCTCCGAAATTCACG 240
QY 241 TCTTCACTAAATTTTCGAAAGAAATCGATCAATACCAACCCATTACACAAAATACATAA 300
DB 241 TCTTCACTAAATTTTCGAAAGAAATCGATCAATACCAACCCATTACACAAAATACATAA 300
QY 301 TCAAAATGGCGAGAATCGTACCTGGAAATTTGCTTCAAGTCGAGAGAGAGAAAGGA 360
DB 301 TCAAAATGGCGAGAATCGTACCTGGAAATTTGCTTCAAGTCGAGAGAGAGAAAGGA 360
QY 361 AGATCGTGGAGAAAGGGGTTAGGGTTAAGCTCAGACTCTTCTAATGAGTAAATGGGACG 420
DB 361 AGATCGTGGAGAAAGGGGTTAGGGTTAAGCTCAGACTCTTCTAATGAGTAAATGGGACG 420
QY 421 GTGTCACATTTTCGGTTTGGAAATGAACCTTTGGGCTCAGCTTATGGGCTATTAGATATT 480
DB 421 GTGTCACATTTTCGGTTTGGAAATGAACCTTTGGGCTCAGCTTATGGGCTATTAGATATT 480
QY 481 TGATGGGCTTTCTAGTAAATAACAATAAAGTTATGGGCTTAGTTTAAATGAAGCCCATGT 540
DB 481 TGATGGGCTTTCTAGTAAATAACAATAAAGTTATGGGCTTAGTTTAAATGAAGCCCATGT 540
QY 541 TGGAAATATTGACACATGCTTGGCTACTAGTCTAAGCAATGCAACCGAACAGTTGTGCG 600
DB 541 TGGAAATATTGACACATGCTTGGCTACTAGTCTAAGCAATGCAACCGAACAGTTGTGCG 600
QY 601 AGACAAGTCGAGCATATACAAATGGATCAAAACAGCCCTAGTGTGCGCGCGTCTGCTCAT 660
DB 601 AGACAAGTCGAGCATATACAAATGGATCAAAACAGCCCTAGTGTGCGCGCGTCTGCTCAT 660
QY 661 GTGTCACTTTGCTTCTCGTTTCTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 720
DB 661 GTGTCACTTTGCTTCTCGTTTCTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 720
QY 721 ATACAAATTTTGGCTGTATCTTGCAAACTCTTCGATCATATCGCAATATACGTGAACA 780
DB 721 ATACAAATTTTGGCTGTATCTTGCAAACTCTTCGATCATATCGCAATATACGTGAACA 780
QY 781 CTGGTGATCTAAATTTGTTGTGTTAAATTTGTTAAATTTAGATTTCTATTCTCCGCTTTAAAG 840
DB 781 CTGGTGATCTAAATTTGTTGTGTTAAATTTGTTAAATTTAGATTTCTATTCTCCGCTTTAAAG 840
QY 841 TGAATTTATATGATCATGTTTAAACATTTGTAAGTAAAGTAAATGATTAATAATGATTAATTT 900
DB 841 TGAATTTATATGATCATGTTTAAACATTTGTAAGTAAAGTAAATGATTAATAATGATTAATTT 900
QY 901 AGTTCATGATAAACGTAAGCAAAATAGATAGATACATTTGATTTTGTGCTATTTTG 960
DB 901 AGTTCATGATAAACGTAAGCAAAATAGATAGATACATTTGATTTTGTGCTATTTTG 960
QY 961 ACATATGCGGAGAGTGAGCTACCGCATGAAGATCAAGAGACATTTGCTCGAGCTCACAG 1020
DB 961 ACATATGCGGAGAGTGAGCTACCGCATGAAGATCAAGAGACATTTGCTCGAGCTCACAG 1020
QY 1021 AGTGAGCTGTAAGAGCTTAGACTGAAGTCCCGCATGCAACCTTAATCTTACGCTGCTCAA 1080
DB 1021 AGTGAGCTGTAAGAGCTTAGACTGAAGTCCCGCATGCAACCTTAATCTTACGCTGCTCAA 1080
QY 1081 ACCAGAGCTCACTTGCACATATATAAATCTCTTAAAGTCCCGTTCTCTTCAATCCATCT 1140
DB 1081 ACCAGAGCTCACTTGCACATATATAAATCTCTTAAAGTCCCGTTCTCTTCAATCCATCT 1140
QY 1141 CTCACAAACAAACAAAAG 1158
DB 1141 CTCACAAACAAACAAAAG 1158
```

```
RESULT 2
US-09-998-059-13
; Sequence 13, Application US/09998059
; Publication No. US20030005485A1
; GENERAL INFORMATION:
; APPLICANT: Ohlrogge, John B.
; APPLICANT: Benning, Christoph
; APPLICANT: Gao, Hongbo
; APPLICANT: Girke, Thomas
; APPLICANT: White, Joseph A.
; TITLE OF INVENTION: Plant Seed Specific Promoters
; FILE REFERENCE: MSU-06689
; CURRENT APPLICATION NUMBER: US/09/998,059
; CURRENT FILING DATE: 2001-11-30
; PRIOR APPLICATION NUMBER: 60/250,401
; PRIOR FILING DATE: 2000-12-01
; NUMBER OF SEQ ID NOS: 60
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 13
; LENGTH: 1164
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; US-09-998-059-13

Query Match 100.0%; Score 1158; DB 10; Length 1164;
Best Local Similarity 100.0%; Pred. No. 1.5e-255;
Matches 1158; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CACAAAATACATACATCAAAATCCAGACTCACATCTACTCAATTTATGCAACTTTCATCATGAA 60
DB 1 CACAAAATACATACATCAAAATCCAGACTCACATCTACTCAATTTATGCAACTTTCATCATGAA 60
QY 61 AACATCAAAACAGTCAAAAGTAACAAATCAAGTCAGATTCAGCACACACAAAGCCAGTAAA 120
DB 61 AACATCAAAACAGTCAAAAGTAACAAATCAAGTCAGATTCAGCACACACAAAGCCAGTAAA 120
QY 121 GATAGAAAATTTAAACGAAACGCTCATGCTAAGCTCGGCAAAATATCTTCCTTAATCAAAACAG 180
DB 121 GATAGAAAATTTAAACGAAACGCTCATGCTAAGCTCGGCAAAATATCTTCCTTAATCAAAACAG 180
QY 181 TAAACAAGAGTAAATAGCAAAATCCGAGCAGAAAGCTCTCACCCACCTCCGAAATTCACG 240
DB 181 TAAACAAGAGTAAATAGCAAAATCCGAGCAGAAAGCTCTCACCCACCTCCGAAATTCACG 240
QY 241 TCTTCACTAAATTTTCGAAAGAAATCGATCAATACCAACCCATTACACAAAATACATAA 300
DB 241 TCTTCACTAAATTTTCGAAAGAAATCGATCAATACCAACCCATTACACAAAATACATAA 300
QY 301 TCAAAATGGCGAGAATCGTACCTGGAAATTTGCTTCAAGTCGAGAGAGAGAAAGGA 360
DB 301 TCAAAATGGCGAGAATCGTACCTGGAAATTTGCTTCAAGTCGAGAGAGAGAAAGGA 360
QY 361 AGATCGTGGAGAAAGGGGTTAGGGTTAAGCTCAGACTCTTCTAATGAGTAAATGGGACG 420
DB 361 AGATCGTGGAGAAAGGGGTTAGGGTTAAGCTCAGACTCTTCTAATGAGTAAATGGGACG 420
QY 421 GTGTCACATTTTCGGTTTGGAAATGAACCTTTGGGCTCAGCTTATGGGCTATTAGATATT 480
DB 421 GTGTCACATTTTCGGTTTGGAAATGAACCTTTGGGCTCAGCTTATGGGCTATTAGATATT 480
QY 481 TGATGGGCTTTCTAGTAAATAACAATAAAGTTATGGGCTTAGTTTAAATGAAGCCCATGT 540
DB 481 TGATGGGCTTTCTAGTAAATAACAATAAAGTTATGGGCTTAGTTTAAATGAAGCCCATGT 540
QY 541 TGGAAATATTGACACATGCTTGGCTACTAGTCTAAGCAATGCAACCGAACAGTTGTGCG 600
DB 541 TGGAAATATTGACACATGCTTGGCTACTAGTCTAAGCAATGCAACCGAACAGTTGTGCG 600
QY 601 AGACAAGTCGAGCATATACAAATGGATCAAAACAGCCCTAGTGTGCGCGCGTCTGCTCAT 660
DB 601 AGACAAGTCGAGCATATACAAATGGATCAAAACAGCCCTAGTGTGCGCGCGTCTGCTCAT 660
QY 661 GTGTCACTTTGCTTCTCGTTTCTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 720
```

Db 661 GTGTGACCTGTTCTCGTGTATTTTAAATTTTTCATAAGTTCTTTGTTTATCTTCA 720
QY 721 ATACAAATTTTGGTGTATCTTGCAGAACTCTTCGATCATATCGCCAAATATACGTGAACA 780
Db 721 ATACAAATTTTGGTGTATCTTGCAGAACTCTTCGATCATATCGCCAAATATACGTGAACA 780
QY 781 CTGTGATCTAAATTTGTTGTTAAATTTAGATTCTATTTCTCCGGTTTAAAG 840
Db 781 CTGTGATCTAAATTTGTTGTTAAATTTAGATTCTATTTCTCCGGTTTAAAG 840
QY 841 TGAATTAATGATCATGTTTAAACATTTGTAAGTAGATGATAATAAATGATAAATTT 900
Db 841 TGAATTAATGATCATGTTTAAACATTTGTAAGTAGATGATAATAAATGATAAATTT 900
QY 901 AGTTGATGATAAAGCTGTAAGCAAAATTTGAGATAGATACATTTGATTTGCTGATTTG 960
Db 901 AGTTGATGATAAAGCTGTAAGCAAAATTTGAGATAGATACATTTGATTTGCTGATTTG 960
QY 961 ACATATCGGAGAGTGAGCTACGCGCATGAAGATCAAGAGACATTTGCTCGAGCTCACAG 1020
Db 961 ACATATCGGAGAGTGAGCTACGCGCATGAAGATCAAGAGACATTTGCTCGAGCTCACAG 1020
QY 1021 AGTACGTTTAAAGCTTAGACTGAACTGCCATGCAAACTAATCCTACGTGGCTCAA 1080
Db 1021 AGTACGTTTAAAGCTTAGACTGAACTGCCATGCAAACTAATCCTACGTGGCTCAA 1080
QY 1081 ACCAGAGCTCACCTTGACAAATATAAATCTCTCTAAGTCCGTTCTCTTATCCATCT 1140
Db 1081 ACCAGAGCTCACCTTGACAAATATAAATCTCTCTAAGTCCGTTCTCTTATCCATCT 1140
QY 1141 CTCACAAACAAACAAAG 1158
Db 1141 CTCACAAACAAACAAAG 1158

RESULT 3

US-09-938-842A-4937
; Sequence 4937, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; TITLE OF INVENTION: SAME, AND METHODS OF USE
; FILE REFERENCE: SRIPI300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 4937
; LENGTH: 1223
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-4937

Query Match 100.0%; Score 1158; DB 9; Length 1223;
Best Local Similarity 100.0%; Pred. No. 1.5e-255;
Matches 1158; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CACAAACATACACTCAAAATCCAGACTCACATCTACTCAATTATGCACTTCATCATGAA 60
Db 63 CACAAACATACACTCAAAATCCAGACTCACATCTACTCAATTATGCACTTCATCATGAA 122
QY 61 ACATCAAAACAGTCAAAAGTAAACAAATCAAGTTCAGATTCAGCACAAAGCCAGTAA 120

Db 123 AACATCAAAACAGTCAAAAGTAAACAAATCAAGTCAGATTTCAGCACAAAGCCAGTAA 182
QY 121 GATAGAAATTTAAACGACGCTCATGCTAAAGCTGCGCAAAATACCTTCTTAATCAAAACAG 180
Db 183 GATAGAAATTTAAACGACGCTCATGCTAAAGCTGCGCAAAATACCTTCTTAATCAAAACAG 242
QY 181 TAAACACAGATTAATTTAGCAAAATCCGAGCAGAAAACCTCTCACCCACCTCCGAAATTCAG 240
Db 243 TAAACACAGATTAATTTAGCAAAATCCGAGCAGAAAACCTCTCACCCACCTCCGAAATTCAG 302
QY 241 TCTTCACATAAAATTTTCGAAAGGATTCGATCAATACCAACCCATTTACAAAAATACATA 300
Db 303 TCTTCACATAAAATTTTCGAAAGGATTCGATCAATACCAACCCATTTACAAAAATACATA 362
QY 301 TCAAAATCGCGAGATTCGATCTGGAACCTTTGCTTCAAGTCGAGAGAGAGAAAAGGA 360
Db 363 TCAAAATCGCGAGATTCGATCTGGAACCTTTGCTTCAAGTCGAGAGAGAGAAAAGGA 422
QY 361 AGATCGTGAGAGAGGGGTTTAAAGCTTCAGACTTCTATTTGGAGTAAATGGGACG 420
Db 423 AGATCGTGAGAGAGGGGTTTAAAGCTTCAGACTTCTATTTGGAGTAAATGGGACG 482
QY 421 GTGTACATTTTCCGTTTGGAAATGAACTTTGGGCTCACGTTATGGGCTATTAGATATT 480
Db 483 GTGTACATTTTCCGTTTGGAAATGAACTTTGGGCTCACGTTATGGGCTATTAGATATT 542
QY 481 TGATGGGCTTTCTAGTAAATACAAATATAAGTTTAAAGCTTCAGACTTCTATTTGGAGTAAATGGGACG 540
Db 543 TGATGGGCTTTCTAGTAAATACAAATATAAGTTTAAAGCTTCAGACTTCTATTTGGAGTAAATGGGACG 602
QY 541 TGGAAATATTGGACATGCTTGGGCTAAGTCTAAACATGCAACCCGAAACAGTTGTCG 600
Db 603 TGGAAATATTGGACATGCTTGGGCTAAGTCTAAACATGCAACCCGAAACAGTTGTCG 662
QY 601 AGACAGTCGCGAGCATATCAATGGATCAACACGCTAGTGTGCGCGCTCTCGCTCAT 660
Db 663 AGACAGTCGCGAGCATATCAATGGATCAACACGCTAGTGTGCGCGCTCTCGCTCAT 722
QY 661 GTGTACCTTTGTTCTCGTTTAAATTTTAAATTTTAAAGTCTCTTTGTTTATCTTCA 720
Db 723 GTGTACCTTTGTTCTCGTTTAAATTTTAAATTTTAAAGTCTCTTTGTTTATCTTCA 782
QY 721 ATACAAATTTTGGCTGTATCTTGGCAACCTCTTCGATCATATCGCCAAATACGTGAACA 780
Db 783 ATACAAATTTTGGCTGTATCTTGGCAACCTCTTCGATCATATCGCCAAATACGTGAACA 842
QY 781 CTGTGATCTAAATTTGTTGTTAAATTTGTTAAATTTAGATTTCTATTCGCGTTTAAAG 840
Db 843 CTGTGATCTAAATTTGTTGTTAAATTTGTTAAATTTAGATTTCTATTCGCGTTTAAAG 902
QY 841 TGAATTAATGATCATGTTTAAACATTTGTAAGTAGATGATAATAAATGATAAATTT 900
Db 903 TGAATTAATGATCATGTTTAAACATTTGTAAGTAGATGATAATAAATGATAAATTT 962
QY 901 AGTTGATGATAAAGCTGTAAGCAAAATTTGAGATAGATAATTGATTTGCTGATTTTG 960
Db 963 AGTTGATGATAAAGCTGTAAGCAAAATTTGAGATAGATAATTGATTTGCTGATTTTG 1022
QY 961 ACATATCGCGAGAGTGAGCTACGCGCATGAAGTCAAGAGACATTTGCTGAGCTCACAG 1020
Db 1023 ACATATCGCGAGAGTGAGCTACGCGCATGAAGTCAAGAGACATTTGCTGAGCTCACAG 1082
QY 1021 AGTGAAGTGAATAAAGCTTAGACTGAAGTCCCATGCAAACTTAATCTACGTGGCTCAA 1080
Db 1083 AGTGAAGTGAATAAAGCTTAGACTGAAGTCCCATGCAAACTTAATCTACGTGGCTCAA 1142
QY 1081 ACAGAGCTCTCTTGACAAATATAAATCTCTCTAAGTCCGTTCTCTTATCCATCT 1140
Db 1143 ACCAGAGCTCACTTGACAAATATAAATCTCTCTAAGTCCGTTCTCTTATCCATCT 1202
QY 1141 CTCACAAACAAACAAAG 1158
Db 1203 CTCACAAACAAACAAAG 1220


```

; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1106
; LENGTH: 6000
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1106

Query Match      4.9%; Score 56.2; DB 15; Length 6000;
Best Local Similarity 51.0%; Pred. No. 0.028;
Matches 133; Conservative 0; Mismatches 128; Indels 0; Gaps 0;

QY 40 ATTATGCAACTTCATCATGAAACATCAAAACAGTCAGTCAAAAGTAAACAAATCAAGTCAAGT 99
Db 4828 ATTAAAAACCAACCGTAAACCCCTTAACGACTTCAACGTAITTTAAAAACAACCTCAAC 4769

QY 100 TCAGCACAAAGCCAGTAAAGATAGAAAAATTTAACGAACGCTCATGCTTAAGCTGCCAA 159
Db 4768 TCAACCTACAAACCCCTACAAAAAACTCTCTAAAATTAITTTCAACTTCAACTCCCTCA 4709

QY 160 AATACTTCTTAATCAAAACAGTAAACGAGTAAATAGCAAAATCCGAGCAGAAAACTCT 219
Db 4708 ATAACTCACTAAATATAAATCTCAAAACCTATAAAAAACCAAAACGTAATCTAACTAAC 4649

QY 220 CACCCACTCGGAATTCACGTCTTCACTAAAATTTTCGAAAGGAATCGATCAATACCA 279
Db 4648 TAAAAACAACCAAAACCGTATCATTTCTTAAATTTAACTACAAACAAAAACCAAA 4589

QY 280 CCCATTACAAAAATACATAA 300
Db 4588 AACATACTACAAAAATAAAA 4568

RESULT 9
US-10-433-793-179
; Sequence 179, Application US/10433793
; Publication No. US20040142334A1
; GENERAL INFORMATION:
; APPLICANT: Epigenomics AG
; TITLE OF INVENTION: Diagnose von mit Angiogenese assoziierten Krankheiten
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/433,793
; CURRENT FILING DATE: 2003-06-06
; NUMBER OF SEQ ID NOS: 212
; SEQ ID NO 173
; LENGTH: 40324
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-433-793-179

Query Match      4.9%; Score 56.2; DB 17; Length 40324;
Best Local Similarity 49.5%; Pred. No. 0.067;
Matches 145; Conservative 0; Mismatches 148; Indels 0; Gaps 0;

QY 669 TTGTTTCCTCGTTTTTTTAAATTTTTCATAAGTCTTTTGTGTTTATCTTCAATACAAAT 728
Db 12605 TTTTTCCTCGTTTTTTTAAATTTTTCATAAGTCTTTTGTGTTTATCTTCAATACAAAT 12664

QY 729 TTTTGGCTGTATCTTGCAAACTCTTCGATCATATCGCAATATACGTCGACACTGGTAT 788
Db 12665 TTTTTCCTCGTTTTTTTAAATTTTTCATAAGTCTTTTGTGTTTATCTTCAATACAAAT 12724

QY 789 CTAAATTTGTTGTTAAATTTGTTAAATTTAGATTTCTATCTCCGGTTTAAAGTGAATAT 848
Db 12725 TTTTTCCTCGTTTTTTTAAATTTTTCATAAGTCTTTTGTGTTTATCTTCAATACAAAT 12784

QY 849 ATGTATCATGTTAAACATTTGTAAGTAAATGATGATAAATAATAGTAAATTTAGTTGATG 908
Db 849 ATGTATCATGTTAAACATTTGTAAGTAAATGATGATAAATAATAGTAAATTTAGTTGATG 908

RESULT 11
US-10-257-166-140
; Sequence 140, Application US/10257166
; Publication No. US20040023230A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Method and Nucleic Acids for Analysing the Methylation of
; FILE REFERENCE: 5013.1011
; CURRENT APPLICATION NUMBER: US/10/257,166
; CURRENT FILING DATE: 2002-10-07

```

```

; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1106
; LENGTH: 6000
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1106

Query Match      4.9%; Score 56.2; DB 15; Length 6000;
Best Local Similarity 51.0%; Pred. No. 0.028;
Matches 133; Conservative 0; Mismatches 128; Indels 0; Gaps 0;

QY 40 ATTATGCAACTTCATCATGAAACATCAAAACAGTCAGTCAAAAGTAAACAAATCAAGTCAAGT 99
Db 4828 ATTAAAAACCAACCGTAAACCCCTTAACGACTTCAACGTAITTTAAAAACAACCTCAAC 4769

QY 100 TCAGCACAAAGCCAGTAAAGATAGAAAAATTTAACGAACGCTCATGCTTAAGCTGCCAA 159
Db 4768 TCAACCTACAAACCCCTACAAAAAACTCTCTAAAATTAITTTCAACTTCAACTCCCTCA 4709

QY 160 AATACTTCTTAATCAAAACAGTAAACGAGTAAATAGCAAAATCCGAGCAGAAAACTCT 219
Db 4708 ATAACTCACTAAATATAAATCTCAAAACCTATAAAAAACCAAAACGTAATCTAACTAAC 4649

QY 220 CACCCACTCGGAATTCACGTCTTCACTAAAATTTTCGAAAGGAATCGATCAATACCA 279
Db 4648 TAAAAACAACCAAAACCGTATCATTTCTTAAATTTAACTACAAACAAAAACCAAA 4589

QY 280 CCCATTACAAAAATACATAA 300
Db 4588 AACATACTACAAAAATAAAA 4568

RESULT 9
US-10-433-793-179
; Sequence 179, Application US/10433793
; Publication No. US20040142334A1
; GENERAL INFORMATION:
; APPLICANT: Epigenomics AG
; TITLE OF INVENTION: Diagnose von mit Angiogenese assoziierten Krankheiten
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/433,793
; CURRENT FILING DATE: 2003-06-06
; NUMBER OF SEQ ID NOS: 212
; SEQ ID NO 173
; LENGTH: 40324
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-433-793-179

Query Match      4.9%; Score 56.2; DB 17; Length 40324;
Best Local Similarity 49.5%; Pred. No. 0.067;
Matches 145; Conservative 0; Mismatches 148; Indels 0; Gaps 0;

QY 669 TTGTTTCCTCGTTTTTTTAAATTTTTCATAAGTCTTTTGTGTTTATCTTCAATACAAAT 728
Db 12605 TTTTTCCTCGTTTTTTTAAATTTTTCATAAGTCTTTTGTGTTTATCTTCAATACAAAT 12664

QY 729 TTTTGGCTGTATCTTGCAAACTCTTCGATCATATCGCAATATACGTCGACACTGGTAT 788
Db 12665 TTTTTCCTCGTTTTTTTAAATTTTTCATAAGTCTTTTGTGTTTATCTTCAATACAAAT 12724

QY 789 CTAAATTTGTTGTTAAATTTGTTAAATTTAGATTTCTATCTCCGGTTTAAAGTGAATAT 848
Db 12725 TTTTTCCTCGTTTTTTTAAATTTTTCATAAGTCTTTTGTGTTTATCTTCAATACAAAT 12784

QY 849 ATGTATCATGTTAAACATTTGTAAGTAAATGATGATAAATAATAGTAAATTTAGTTGATG 908
Db 849 ATGTATCATGTTAAACATTTGTAAGTAAATGATGATAAATAATAGTAAATTTAGTTGATG 908

```


Db	925	AAACCACTATATAAAAAACAAAAATAAAATTCCTCAAAAAAATAAAATCAAAACGATAATA	866
QY	196	AGCAAAATCCGAGCAGAAAACTCTCACCCACCTCCGAAATTCACGCTTCCTCAATAATTT	255
Db	865	ATCTAACAAATCTTTATAATATATATCTCAAAAAAATAAATCAATACATTAATAAATA	806
QY	256	TGGAAGGAATCGATCAATACCAACCCCAATTACAAAAATACATTAATCAAAA	306
Db	805	TCTACACTCCCAATTTTATTACAAACATATTCAACAATCAATAATAATAA	755

RESULT 14

```

RESULTS 14
US-10-240-485-79/c
; Sequence 79, Application US/10240485
; Publication No. US20030148327A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with
; TITLE OF INVENTION: Metastasis
; FILE REFERENCE: 5013.1007
; CURRENT APPLICATION NUMBER: US/10/240,485
; CURRENT FILING DATE: 2002-10-02
; PRIOR APPLICATION NUMBER: PCT/EP01/03970
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 202
; SEQ ID NO 79
; LENGTH: 6713
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-240-485-79

```

RESULT 15

RESULT IS
US-10-021-323-7699/c
; Sequence 7699, Application US/10021323
; Publication No. US20040123340A1
; GENERAL INFORMATION:

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 8, 2004, 23:43:11 ; Search time 5779 seconds
(without alignments)
7301.819 Million cell updates/sec

Title: US-09-998-059-1
Perfect score: 1158
Sequence: 1 cacaaacatacactcaaaat.....ctctcacacacacacaaaag 1158

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 32822875 seqs, 18219865908 residues

Total number of hits satisfying chosen parameters: 65645750

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: gb_est1:*

2: gb_est2:*

3: gb_hic:*

4: gb_est3:*

5: gb_est4:*

6: gb_est5:*

7: gb_est6:*

8: gb_gss1:*

9: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	395.4	34.1	413	8	BH911706	BH911706 SALK_0716
C 2	333.4	28.8	400	8	BZ289690	BZ289690 SALK_0230
C 3	145.2	12.5	692	8	BZ084475	BZ084475 l1b61d08.
C 4	129.4	11.2	650	8	BH555922	BH555922 BOHX116TR
C 5	83	7.2	963	9	CLS15027	CLS15027 SAIL_895
C 6	64	5.5	987	9	CNS014FQ	AL104456 Drosophil
C 7	61.6	5.3	1101	9	CNS017V2	AL108536 Drosophil
C 8	60.6	5.2	928	9	CNS00DKY	AL071865 Drosophil
C 9	60.4	5.2	1101	9	CNS0016L1	AL06896 Drosophil
C 10	60.2	5.2	922	9	CNS0073W	AL066784 Drosophil
C 11	60	5.2	759	9	CNS006QXV	AL411257 T7 end of
C 12	59.4	5.1	324	9	CNS0243U	AL180291 Tetraodon
C 13	59.4	5.1	1187	9	CL648142	CL648142 CH213-164
C 14	59.2	5.1	1092	9	CNS020K7	AL175696 Tetraodon
C 15	58.8	5.1	1045	9	CNS03YF4	AL266197 Tetraodon
C 16	58.4	5.0	1001	9	CNS001400	AL103554 Drosophil
C 17	57.8	5.0	1101	9	CNS000L72	AL078714 Drosophil
C 18	57.6	5.0	627	1	A1525005	A1525005 promna-4
C 19	57	4.9	873	9	CNS024M5	AL180950 Tetraodon
C 20	57	4.9	987	9	CNS014FQ	AL104456 Drosophil
C 21	57	4.9	1039	9	CNS006H7	AL064351 Drosophil
C 22	57	4.9	1146	9	CNS021G2	AL176843 Tetraodon
C 23	56.8	4.9	555	5	BQ739452	BQ739452 PfESToab4
C 24	56.8	4.9	583	9	CNS044QG	AL274417 Tetraodon

C 25	56.6	4.9	924	9	CNS07A5L	AL436159 T7 end of
C 26	56.6	4.9	982	9	CNS004WF	AL055517 Drosophil
C 27	56.4	4.9	764	8	CF289381	CF289381 AGENCOURT
C 28	56.4	4.9	886	8	BH177277	BH177277 008_L_22-
C 29	56.4	4.9	886	9	CNS07JUX	AL614235 T3 end of
C 30	56.4	4.9	966	9	CNS006R7	AL065822 Drosophil
C 31	56.2	4.9	892	8	AZ531076	AZ531076 ENTHX19TF
C 32	56.2	4.9	911	8	AZ691769	AZ691769 ENTHX19TF
C 33	56	4.8	693	9	CNS006MN	AL065759 Drosophil
C 34	56	4.8	1001	9	CNS0155H	AL105023 Drosophil
C 35	55.8	4.8	892	9	CL101621	CL101621 ISB1-38N5
C 36	55.8	4.8	1101	9	CNS001FB	AL060732 Drosophil
C 37	55.8	4.8	2300	3	CR734152	CR734152 Tetraodon
C 38	55.6	4.8	920	9	CNS0062R	AL061710 Drosophil
C 39	55.4	4.8	843	9	CNS000CS1	AL059666 Drosophil
C 40	55.2	4.8	839	9	AGS18301	AGS18301 Mus muscu
C 41	55.2	4.8	928	9	CNS000DKY	AL071865 Drosophil
C 42	55.2	4.8	1101	9	CNS000LT2	AL078714 Drosophil
C 43	55.2	4.8	1201	9	CNS0168L	AL106431 Drosophil
C 44	55	4.7	888	9	CNS007VQ	AL050896 Drosophil
C 45	54.6	4.7	1092	9	CNS020K7	AL175696 Tetraodon

ALIGNMENTS

RESULT 1
BH911706/c

LOCUS

DEFINITION

Arabidopsis thaliana genomic clone SALK_071668.55.25.x, genomic survey sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Class: TDNA tagged.

Location/Qualifiers

1. 413

/organism="Arabidopsis thaliana"

/mol_type="genomic DNA"

/ecotype="Col-0"

/db_xref="taxon:3702"

/clone="SALK_071668.55.25.x"

/note="PCR was performed on Arabidopsis thaliana lines each of which contains one or more TDNA insertion elements. The resultant fragment for each line was directly sequenced to determine the genomic sequence at the site of insertion. Details of the protocols used can be found at http://signal.salk.edu/tdna_protocols.html"

ORIGIN

source

features

features

features

features

features

features

features

features

features

features

features

features

features

features

features

features

features

features

features

features

each of which contains one or more TDNA insertion elements. The resultant fragment for each line was directly sequenced to determine the genomic sequence at the site of insertion. Details of the protocols used can be found at http://signal.salk.edu/tdna_protocols.html

ORIGIN		Query Match	28.8%;	Score 333.4;	DB 8;	Length 400;
		Best Local Similarity	94.1%;	Pred. No. 5.7e-65;		
		Mismatches	0;	Mismatches	21;	Indels
		Matches	368;	Conservative	0;	Gaps
QY	3	CAACATACACATCAAAATCCAGACTCACATCTACTCAATTAATGCAATTCATCATCAAAA	62			
Db	413	CAACATACACTCAAAATCCAGACTCACATCTACTCAATTAATGCAATTCATCATCAAAA	354			
QY	63	CATCAAAAACAGTCAAAAGTAACAAAATCAAGTCAGATTCAGCACACAAAAGCCAGTAAAGA	122			
Db	353	CATCAAAAACAGTCAAAAGTAACAAAATCAAGTCAGATTCAGCACACAAAAGCCAGTAAAGA	294			
QY	123	TAGAAAATTAAACGACGCTCATGCTTAAGTGGCGAAAATACTTCTTAATCAAAAACAGTA	182			
Db	293	TAGAAAATTAAACGACGCTCATGCTTAAGTGGCGAAAATACTTCTTAATCAAAAACAGTA	234			
QY	183	ACACAGGTAATTAAGCAAAATCCGACGAGAAAATCTCCACCCACTCCGAAAATTCAGTCT	242			
Db	233	ACACAGGTAATTAAGCAAAATCCGACGAGAAAATCTCCACCCACTCCGAAAATTCAGTCT	174			
QY	243	TTCACTAAAAATTTTCGAAAGGAATCGATCAATACCAACCCATTACACAAAATACATAATC	302			
Db	173	TTCACTAAAAATTTTCGAAAGGAATCGATCAATACCAACCCATTACACAAAATACATAATC	114			
QY	303	AAAATGGCGAGATCGTACCTGGAACTTTGCTTCAAGTCGACGAGAGAGAAAAGGAAG	362			
Db	113	AAAATGGCGAGATCGTACCTGGAACTTTGCTTCAAGTCGACGAGAGAGAAAAGGAAGTA	54			
QY	363	ATCGTCGAGAAAGGGTTTAAAGCTCAGACTTCTATTGGAGTAAATG	415			
Db	53	GATTCGGAGAAAGGGTTTAAAGCTCAGACTTCTATTGGAGTAAATG	1			

RESULT 2	BZ289690	400 bp	DNA	linear	GSS 24-OCT-2002
LOCUS	SALK_023089.34.95.x Arabidopsis thaliana	TDNA insertion lines			
DEFINITION	Arabidopsis thaliana genomic clone SALK_023089.34.95.x, genomic survey sequence.				
ACCESSION	BZ289690				
VERSION	BZ289690.1	GI:24331741			
KEYWORDS	GSS.				
SOURCE	Arabidopsis thaliana (thale cress)				
ORGANISM	Arabidopsis thaliana				
REFERENCE	Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R., Gadriab,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L., Shinn,P., Zimmermann,J. and Ecker,J.R.				
AUTHORS	A Sequence-Indexed Library of Insertion Mutations in the Arabidopsis Genome				
TITLE	Unpublished (2001)				
JOURNAL	Contact: Joseph R. Ecker				
COMMENT	Salk Institute Genomic Analysis Laboratory (SIGAL) The Salk Institute for Biological Studies 10010 N. Torrey Pines Road, La Jolla, CA 92037, USA Tel: 858 453 4100 x1752 Fax: 858 558 6379 Email: eckersalk.edu This is single pass sequence recovered from the left border of TDNA.				
FEATURES	Class: TDNA tagged.				
source	Location/Qualifiers				
	1..400				
	/organism="Arabidopsis thaliana"				
	/mol_type="genomic DNA"				
	/ecotype="Col-0"				
	/db_xref="taxon:3702"				
	/clone="SALK_023089.34.95.x"				
	/clone_lib="Arabidopsis thaliana TDNA insertion lines"				
	/note="PCR was performed on Arabidopsis thaliana lines"				

RESULT 3	BZ084475	692 bp	DNA	linear	GSS 10-OCT-2002
LOCUS	11b6ld08_g1 B.oleracea002	Brassica oleracea	genomic,	genomic survey	
DEFINITION	sequence.				
ACCESSION	BZ084475				
VERSION	BZ084475.1	GI:23714246			
KEYWORDS	GSS.				
SOURCE	Brassica oleracea				
ORGANISM	Brassica oleracea				
REFERENCE	Eukaryota, Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Brassica.				
AUTHORS	1 (bases 1 to 692) Delehaunty,K., Fewell,G., Fulton,L., McCombie,W.R., Miner,T., Nash,W., Rabinowicz,P.D. and Wilson,R.K.				
TITLE	Whole genome shotgun reads from Brassica oleracea				
JOURNAL	Unpublished (2002)				
COMMENT	Contact: Richard K. Wilson Genome Sequencing Center Washington University School of Medicine Email: submissions@wustl.edu Plate: 11b61 row: d column: 08 Seq primer: -28RPOT reverse Class: shotgun High quality sequence start: 196 High quality sequence stop: 551. Location/Qualifiers				
FEATURES	1..692				
source	/organism="Brassica oleracea"				
	/mol_type="genomic DNA"				
	/db_xref="taxon:3712"				
	/clone_lib="B.oleracea002"				

/note="Vector: pOTw13; Whole genome shotgun library from flowering buds. DNA was purified from a crude nuclear prep using Brassica oleracea T01000DH3 buds provided by Thomas Osborn at the University of Wisconsin. Genomic DNA was provided by Pablo Rabinowicz (CSHL) and the shotgun library prepared at Washington University Genome Sequencing Center."

ORIGIN

Query Match 12.5%; Score 145.2; DB 8; Length 692;
Best Local Similarity 69.8%; Pred. No. 2.8e-22;
Matches 268; Conservative 0; Mismatches 108; Indels 8; Gaps 5;

Qy 770 ATACGTGAACACTGGTGTATCTAAATTTGTTGTTAAATTTGT-TAAATTTAGATTTCTATTCT 828
|||
Db 195 ATAAAGGGATACCTAGTGTATGTAATTTGATACGTTAAATTTGTCAAAAAGTAGATTTCTCTAT 254
|||
Qy 829 CCGGTTTAAAGTGAATTAATATGATCATGTTTAAACATTTGTAAGTAAGATGATAATAA 888
|||
Db 255 CCCCTTTATTATATGTTTAAATGAATTTAAG---ACGATTAAAAATGAATTAATCAACT 311
|||
Qy 889 AATGATAAAATTTAGTGTGATGATAACGTGAGCAAAAATGAGATAGATACATTTGATTT 948
|||
Db 312 AGAGTTCTTGTACTGTTGATTAACATGAACCTAAAAAATGAGAGATACAT-TTGAATTT 370
|||
Qy 949 TGTGTAATTTTGACATATGCGGAGAGTGAGCTACGCGCATGAAGATCAAGAGACACTTGC 1008
|||
Db 371 TGTGCTGTTTGGAGTATGCGGAGATAGTGTCTACGCGCATGAACATCATGAGACACTTGC 430
|||
Qy 1009 TCGAGCTACAGAGTAGCTGTGTAAGAGCTTTAGACTGAAGTCCCAACCTAATCC 1068
|||
Db 431 TTCAAGCTACAGAGTAGCTGTGTAAGAGCTTTAGACTGAAGTCCCAACCTAATCC 490
|||
Qy 1069 TAGCTGCTCAAAACCGAGCTCCTTGTGACAAATATATAAACTCTCTTAAGTCCCGTTC- 1127
|||
Db 491 TAGTGGCAAAACCTTCATGCTCACTCCAC-ATATATAAACTCTTACCAAGTCTCCATCT 549
|||
Qy 1128 -TCTTCATCCATCTCTCACAAACAA 1150
|||
Db 550 TTCTTCATCCATCTATCACAAAAA 573
|||

RESULT 4
BH655922 650 bp DNA linear GSS 19-FEB-2002
LOCUS BOHX116R BO 2.3 KB Brassica oleracea genomic clone BOHX116,
DEFINITION genomic survey sequence.
ACCESSION BH655922
VERSION BH655922.1 GI:18714215
KEYWORDS GSS.
SOURCE Brassica oleracea
ORGANISM Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
1 (bases 1 to 650)
Town, C.D., Van Aken, S., Utterback, T., Koo, H. and Fraser, C.M.
TITLE Whole genome shotgun sequencing of Brassica oleracea
JOURNAL Unpublished (2001)
COMMENT Contact: Chris Town

TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TR
Class: sheared ends.

FEATURES
source

Location/Qualifiers
1..650
/organism="Brassica oleracea"
/mol_type="genomic DNA"
/strain="T01000DH3"

ORIGIN

Query Match 11.2%; Score 129.4; DB 8; Length 650;
Best Local Similarity 77.9%; Pred. No. 1.1e-18;
Matches 194; Conservative 0; Mismatches 51; Indels 4; Gaps 3;

Qy 904 TGATGGATAACCTGAACCAAAAATGAGATAGATACATTTGATTTGCGTATTTTGACA 963
|||
Db 9 TGGTTGATAACATGAATCTTAAAAATGAGATACA-TTTGAATTTTGTGCGTGGTTGGAG 67
|||
Qy 964 TATCGGAGAGTGAGCTACGCGCATGAAGATCAAGAGACACTTCTCGAGCTCACAGAT 1023
|||
Db 68 TATCGGAGAGTAGTGTCTACGCGCATGAACATCATGAGACACTTCTTCAGCTCACAGAT 127
|||
Qy 1024 GACGTGTAAAAAGCTTAGACTGAAAGTCCCATGCAAAACCTTAATCTAGTGGCTCAAAAC 1083
|||
Db 128 GACGTGTAAAAAGCATAGACCCACGACTTCATGCAAAACCTTCTAGTGGCAAAACCT 187
|||
Qy 1084 ACGAGCTCACTTGACAATATATAAACTCCCTTAAGTCCCGTTC--TCTTCATCCATCTC 1141
|||
Db 188 TCATGCTCACTCCAC-ATATATAAACTCTTACCAAGTCTCCATCTTCTTCATCCATCTA 246
|||
Qy 1142 TCACAACAA 1150
|||
Db 247 TCACAANA 255
|||

RESULT 5
CL515027 963 bp DNA linear GSS 01-APR-2004
LOCUS SAIL_895_F04.v1 SAIL Collection Arabidopsis thaliana genomic clone
DEFINITION SAIL_895_F04.v1, genomic survey sequence.
ACCESSION CL515027
VERSION CL515027.1 GI:46012347
KEYWORDS GSS.
SOURCE Arabidopsis thaliana (thale cress)

ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 963)
Sessions, A., Burke, E., Presting, G., Aux, G., McElver, J., Patton, D.,
Dietrich, B., Ho, P., Bacwaden, J., Ko, C., Clarke, J.D., Cotton, D.,
Bullis, D., Snell, J., Miguel, T., Hutchison, D., Kimmerly, B.,
Mitzel, T., Katagiri, F., Glazebrook, J., Law, M. and Goff, S.A.
TITLE A high-throughput Arabidopsis reverse genetics system
JOURNAL Plant Cell 14 (12), 2985-2994 (2002)
MEDLINE 22356987
PUBMED 12468722
COMMENT Contact: Sessions A

REFERENCE
AUTHORS Applied Trait Genetics
Syngenta Biotechnology Inc.
3054 Cornwalis Rd., Research Triangle Park, NC 27709, USA
Email: allen.sessions@syngenta.com
ABRC Stock Number CS840399; T-DNA left border flanking sequences of
Syngenta Arabidopsis Insertion Library (SAIL) lines are available
through the Arabidopsis Biological Resource Center (ABRC).
Sequences represent a pool of amplified genomic regions and not
single contiguous sequences.

Class: TDNA tagged.
Location/Qualifiers
1..963
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/ecotype="Columbia"
/db_xref="taxon:3702"
/clone="SAIL_895_F04.v1"
/clone_lib="SAIL Collection"

FEATURES
source

Location/Qualifiers
1..963
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/ecotype="Columbia"
/db_xref="taxon:3702"
/clone="SAIL_895_F04.v1"
/clone_lib="SAIL Collection"

/db_xref="taxon:3712"
/clone="BOHX116"
/clone_lib="BO 2.3 KB"
/note="Vector: pHO51; Site 1: BstXI; 2-3 kb sheared
genomic DNA inserted into pHO51 using BstXI linkers"

• D'h

Query Match

ORIGIN

[illegible]

RESULT 8	
CNS000DKY/c	
LOCUS	
DEFINITION	CNS000DKY 928 bp DNA linear GSS 04-JUN-1999 Drosophila melanogaster genome survey sequence T7 end of BAC # BACR27A24 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION	AL071865
VERSION	AL071865.1
KEYWORDS	GI:4948170
SOURCE	GSS.
ORGANISM	Drosophila melanogaster (fruit fly) Drosophila melanogaster Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila. 1 (bases 1 to 928)
REFERENCE	

AUTHORS	TITLE	JOURNAL	COMMENT
Genoscope. Direct Submission Submitted (02-JUN-1999) BP 191 91006 ERYV cedex - FRANCE - Web : www.genoscope.cns.fr - E-mail : seqref@genoscope.cns.fr	Genoscope - Centre National de Sequencage : BP 191 91006 ERYV cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)		Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Oosagawa and Aaron Mammeter in Piter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPci-98 and was constructed by partial, EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's p1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila.bac.htm .

```

FEATURES
    source
        bound at http://bacpac.med.buffalo.edu/drosophila\_bac.htm.
        Location/Qualifiers
            1..928
                /organism="Drosophila melanogaster"
                /mol_type="genomic DNA"
                /db_xref="taxon:7227"
                /clone="BACR27A24"
                /clone_lib="RPCI-98"
                /note="end = 77"

```

[illegible][illegible]

RESULT 9
CNS016LT

[illegible]

DEFINITION

ACCESSION

VERSION
KEYWORDS

KEYWORDS
SOURCE
ORGANISM

ORGANISM

REFERENCE AUTHORS

THE JOURNAL OF THE

JOURNAL

COMMENT

FEATURES

amos

ORIGIN

Query Mat.

Best Local Matches

609

○ ○

○ ○

71

Qy 7:

8. Db

Qy 756

906 qb

85
Cy

900

96

912y

	Matches	132;	Conservative	0;	Mismatches	125;	Indels	0;	Gaps	0;
QY	657	TCATGTGTCACCTGTTCTCGTCTTTTTTTTTTAAATTTTCATAAGTCTCTTTGTTTTATC	716							
Db	896	TAAATTTTAATGATTTTATAATTTATTTTATTTTGAATAATATTTTTATTATT	955							
QY	717	TTCAATACAAATTTTTGGCTGTACTTGCAAACTCTCGATCATATCGCCAATATACGG	776							
Db	956	TTATTATATTTTATATATTTTTTTTTTATATGTTTTATATTTAAATTTNATTATATATT	1015							
QY	777	AACACTGGTGAICTAATTTCTGTGTTAAATTTAGATCTATTCTCCGGTTTA	836							
Db	1016	AACTTTATGAATATTTTGGTTATTTTTTTTAAATTTTATAAAAATATATTTTTTTA	1075							
QY	837	AAAGTCAATATATGTATCATGCTTAAAAATTGTGAAGTAAATGATAATAAATGATAA	896							
Db	1076	TAAGTATTTATTTAAATAATGTATTATATTTTTTAAATAATGTTTATAATNATATAT	1135							
QY	897	ATTAGTTGATGATAA	913							
Db	1136	TTNTATNTATTTATTAA	1152							

RESULT 14	CNS020K7/c	1092 bp	DNA	linear	GSS 01-SEP-2000
LOCUS					
DEFINITION	Tetraodon nigroviridis genome survey sequence T7 end of clone 222L11 of library G from Tetraodon nigroviridis, genomic survey sequence.				

ACCESSION	AL175696
VERSION	AL175696.1 GI:7813753
KEYWORDS	GSS; genome survey sequence.
SOURCE	Tetraodon nigroviridis
ORGANISM	Tetraodon nigroviridis
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodon.
REFERENCE	1
AUTHORS	Roest Croollux H., Jaillon O., Dasilva C., Bouneau L., Fisher C., Bernot A., Pizanes C., Wincker P., Brottier P., Quetier P., Saurin W. and Weissenbach J.
TITLE	Estimate of human gene number provided by genome-wide analysis using Tetraodon nigroviridis DNA sequence
JOURNAL	Nat. Genet. 25 (2), 235-238 (2000)
MEDLINE	20296633
PUBMED	10835645

REFERENCE	2	Roest Crollius, H., Jaillon, O., Dasilva, C., Ozouf-Costaz, C., Pizanes, C., Fischer, C., Bouneau, L., Billault, A., Quetier, F., Saurin, W., Bernot, A. and Weissenbach, J.
AUTHORS		Characterization and repeat analysis of the compact genome of the freshwater pufterfish Tetraodon nigroviridis
TITLE		Genome Res. 10 (7), 939-949 (2000)
JOURNAL		
MEDLINE		20359837
PUBMED		10899143
REFERENCE	3	(bases 1 to 1092)
AUTHORS		Genoscope.
TITLE		Direct Submision
JOURNAL		Submitted (12-APR-2000) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT		This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at http://www.genoscope.cns.fr/Tetraodon .

```

FEATURES
    source
        1. i1092
            /organism="Tetraodon nigroviridis"
            /mol_type="genomic DNA"
            /db_xref="taxon:99883"
            /clone="222111"
            /clone_lib="G"
            /location/Qualifiers
                url="http://www.genoscope.cns.fr/terramoon"

```

						/note="Genoscope sequence ID : C0AG222C0F06LPI-end : T7"							
	Query Match	5.1%;	Score 59.2;	DB 9;	Length 1092;								
	Best Local Similarity	39.9%;	Pred. No. 0.009;										
	Matches 123;	Conservative 47;	Mismatches 137;	Indels 1;	Gaps 1;								
<hr/>													
Qy	659	ATGTGTCACCTGTTTCCCTCGTTTTTTTAAATTTTTCATAAGCTCTTTTGTTTATCTT	718										
Dd	1025	ATATTATTATTTTATATTATTTTTTTTTTTTTTAAAAATAAATAAAMTTTTNMT	966										
<hr/>													
Qy	719	CAATACAAATTTTTGGCTGTATCTGC AAACTCTTCGATCATATCGCCAATATACGTGA	778										
Dd	965	TTATATTTTTTTTTTTTTTWT -TAWAAAAAAAAAAAAAAAAAAWATTAWAWTWAWAA	907										
<hr/>													
Qy	779	CACGTGGTAGCTAAATTTGTTGTGTAAATTTAGATTCCTATCTCCGGTTTAAA	838										
Dd	906	AAAAAAWWMTTWTTTTTTAATTTTTTTTTTTTAWATTWAAATTTWAAAATWTTTTAAAW	847										
<hr/>													
Qy	839	AGTCAAATTATATGTATCATGTTTAAAAACATTTGAAGTAAGATGATAATAAAATGATAAT	898										
Dd	846	AATTATTITTTTWTTTTATWATAWTAANAATNTTTTTTTTAAAAATAWAANWTATWWWATAAA	787										
<hr/>													
Qy	899	TTAGTTGATGATCAACGTGAAGCAAAAAATGAGATAGATACATTTGATTTTGTGCGTATTT	958										
Dd	786	WTAAWAWWTTTWTTTTTTWWWWWWTAAABAAWAAKTWWTTTWTTTTTTTTWTWTWTAT	727										
<hr/>													
Qy	959	TGACATAT	966										
Dd	726	TWATTTTT	719										

RESULT 15
CNS03YE4/c
LOCUS
DEFINITION
Tetraodon nigroviridis genome survey sequence T7 end of clone
068003 of library G from Tetraodon nigroviridis, genomic survey
sequence.
ACCESSION
AL266197
VERSION
AL266197.1 GI:7987962
KEYWORDS
GSS; genome survey sequence.
SOURCE
Tetraodon nigroviridis
Tetraodon nigroviridis
Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Actinomorpha; Acanthopterygii; Percormorpha; Tetraodontiformes;
Tetraodontidae; Tetraodontidae; Tetraodon.

1	Roest Crolius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F., Saurin,W. and Weissenbach,J. Estimate of human gene number provided by genome-wide analysis using Tetraodon nigroviridis DNA sequence Nat. Genet. 25 (2), 235-238 (2000) 20296633 10835645	Tetraodonbouneau, Tetraodoncontractus
2	Roest Crolius,H., Jaillon,O., Dasilva,C., Ozouf-Costaz,C., Fizames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and Weissenbach,J. Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis Genome Res. 10 (7), 939-949 (2000) 20359837 10899143	
3	(bases 1 to 1045) Genoscope. Direct Submission Submitted (12-APR-2000) Genoscope - Centre National de Sequencage BP 191 91006 EVRY cedex - FRANCE [E-mail : seqref@genoscope.cns.fr] - Web : www.genoscope.cns.fr)	
4	This sequence is a single read and was generated as part of a large scale sequencing project of the Tetraodon nigroviridis	

COMMENT

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 8, 2004, 17:32:55 ; Search time 866 Seconds

(without alignments)

7019.434 Million cell updates/sec

Title: US-09-998-059-1

Perfect score: 1158

Sequence: 1 cacaacatacactcaaat.....ctctcacaacaacaaaag 1158

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 4134886 seqs, 2624710521 residues

Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_23Sep04:*

1: Geneseqn1980s:*

2: Geneseqn1990s:*

3: Geneseqn2000s:*

4: Geneseqn2001s:*

5: Geneseqn2001bs:*

6: Geneseqn2002s:*

7: Geneseqn2002bs:*

8: Geneseqn2003s:*

9: Geneseqn2003bs:*

10: Geneseqn2003cs:*

11: Geneseqn2003ds:*

12: Geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1158	100.0	1164	10	ADJ19001	Adj19001 Thale cre
2	1158	100.0	1223	6	ABZ17132	Abz17132 Arabidops
C 3	81	7.0	1118	3	AAC33222	Aac33222 Arabidops
C 4	65.4	5.6	10326	6	ABL34166	Ab134166 Human imm
C 5	56.2	4.9	6000	6	ABL33133	Ab133133 Human imm
C 6	56.2	4.9	40324	6	ABQ67149	Abq67149 Human ang
7	55.8	4.8	5845	6	ABL33662	Ab133662 Human imm
8	54.6	4.7	7441	6	ABK40058	Abk40058 Human che
9	54.6	4.7	7479	6	AAS63345	Aas63345 Chemicall
C 10	54.2	4.7	6713	6	ABL33080	Ab133080 Human imm
C 11	54.2	4.7	6713	6	ABL34526	Ab134526 Human met
C 12	54.2	4.7	6713	6	ABL70251	Ab170251 Chemicall
C 13	53.6	4.6	7215	6	ABN91164	Abn91164 Staphyloc
C 14	53.4	4.6	29993	10	ADB37662	Adb37662 Human che
15	52.4	4.5	2000	8	ADA71938	Ada71938 Rice gene
16	52.2	4.5	337	8	ABX41811	Abx41811 Bovine ES
17	52	4.5	694	4	AAS53104	Aah53104 S. epider
18	52	4.5	875	4	AAI95044	Aai95044 Human neu
19	52	4.5	3028	4	AAH54496	Aah54496 S. epider
20	52	4.5	3032	4	AAH54863	Aah54863 S. epider
21	52	4.5	6968	4	AAH52683	Aah52683 S. epider

22	51.2	4.4	96588	9	ADA03026	Ada03026 Human MBN
23	51.2	4.4	96588	10	ADB72764	Adb72764 Human MBN
24	51.2	4.4	96588	10	ADC85506	Adc85506 Human Mbn
25	51.2	4.4	96588	12	ADM74621	Adm74621 Human car
C 26	51	4.4	11735	4	AAS45328	Aas45328 Chemicall
C 27	51	4.4	11735	6	ABK28167	Abk28167 DNA trans
C 28	51	4.4	11735	6	AAS61141	Aas61141 Human gen
C 29	51	4.4	13511	6	ABL32280	Ab132280 Human imm
C 30	50.8	4.4	6621	8	ABZ10129	Abz10129 Haematopo
C 31	50.8	4.4	6621	8	ABZ10243	Abz10243 Haematopo
C 32	50.8	4.4	6621	10	ADB84153	Ade84153 Human lym
C 33	50.6	4.4	3683	8	ABZ10199	Abz10199 Haematopo
C 34	50.6	4.4	8011	6	ABL32078	Ab132078 Human imm
35	50.6	4.4	8011	6	AD28368	Ad28368 Human che
36	50.4	4.4	419	8	ABX46069	Abx46069 Bovine ES
37	50.4	4.4	7544	4	AAS45300	Aas45300 Chemicall
38	50.4	4.4	7544	6	ABK28139	Abk28139 DNA trans
39	50.2	4.3	7037	6	ABL33751	Ab133751 Human imm
40	50.2	4.3	13503	6	ABL34070	Ab134070 Human imm
41	50	4.3	597	6	ABQ36487	Abq36487 Oligonuc1
C 42	50	4.3	597	6	ABQ36486	Abq36486 Oligonuc1
43	50	4.3	5241	6	ABL70492	Ab170492 Chemicall
44	50	4.3	5241	6	AAS61450	Aas61450 Human gen
C 45	49.8	4.3	556	5	ABV40163	Abv40163 Human pro

ALIGNMENTS

RESULT 1
ADJ19001
ID ADJ19001 standard; DNA; 1164 BP.
XX
AC ADJ19001;
XX
DT 20-MAY-2004 (first entry)
XX
DE Thale cress seed-specific promoter region Pl DNA.
XX
KW plant; seed-specific promoter region; seed development; food;
KW agricultural; thale cress; ds; Pl.
XX
OS Arabidopsis thaliana.
XX
PN US2003005485-A1.
XX
PD 02-JAN-2003.
XX
PF 30-NOV-2001; 2001US-00998059.
XX
PR 01-DEC-2000; 2000US-0250401P.
XX
(OHLR/) OHLROGGE J B.
PA (BENN/) BENNING C.
PA (GAOH/) GAO H.
PA (GIRK/) GIRKE T A A.
PA (WHIT/) WHITE J A.
XX
PI Ohlrogge JB, Benning C, Gao H, Girke TAA, White JA;
XX WPI; 2003-370848/35.
XX
New DNA comprising plant seed specific promoters, useful for controlling expression of genes in plants, particularly for modifying seed products (proteins, carbohydrates or oils), which are of major economic and food values.
XX
PS Claim 1; SEQ ID NO 1; 57pp; English.
XX
CC The invention relates to a novel isolated DNA molecule which comprises a plant seed-specific promoter region. The DNA molecule of the invention may be useful for controlling the expression of genes in plants during the different phases of seed development, in particular, for modifying

CC seed products, which are the major economic and food values of most
 CC agricultural plants e.g. as sources of proteins, carbohydrates or oils.
 CC The current sequence is that of the thale cress seed-specific promoter
 CC region Pl DNA of the invention.
 XX
 SQ Sequence 1164 BP; 399 A; 228 C; 205 G; 332 T; 0 U; 0 Other;
 Query Match 100.0%; Score 1158; DB 10; Length 1164;
 Best Local Similarity 100.0%; Pred. No. 4.6e-250;
 Matches 1158; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 CACAAACATACACTCAAAATCCGAGCTCAGATCTACTCTCAATATGCACTTCATCATGAA 60
 DB 1 CACAAACATACACTCAAAATCCGAGCTCAGATCTACTCTCAATATGCACTTCATCATGAA 60
 QY 61 AACATCAAAACAGTCAAAAGTAAACAAATCAAGTCAAGTTCAGACACACAAAGCCAGTAAA 120
 DB 61 AACATCAAAACAGTCAAAAGTAAACAAATCAAGTTCAGACACACAAAGCCAGTAAA 120
 QY 121 GATAGAAAATTTAAACGAGCTCATGCTAAGCTGCGCAAAATCTTCTTAATCAAAACAG 180
 DB 121 GATAGAAAATTTAAACGAGCTCATGCTAAGCTGCGCAAAATCTTCTTAATCAAAACAG 180
 QY 181 TAACACGAGTAATTTAGCAAAATCCGAGAGAGAAATCTCAACCACTCCGAAATTCACG 240
 DB 181 TAACACGAGTAATTTAGCAAAATCCGAGAGAGAAATCTCAACCACTCCGAAATTCACG 240
 QY 241 TCTTCACTAAAATTTTGAAGGAATCGATCAATCAACCCCATTAACAAAATACATAA 300
 DB 241 TCTTCACTAAAATTTTGAAGGAATCGATCAATCAACCCCATTAACAAAATACATAA 300
 QY 301 TCAAAATGGCGAGAAATCGTACCTGGAACTTTGGTTCAGTTCGAGAGAGAAAGGA 360
 DB 301 TCAAAATGGCGAGAAATCGTACCTGGAACTTTGGTTCAGTTCGAGAGAGAAAGGA 360
 QY 361 AGATCGTGGAGAAAGGGTTTAGGGTTTAAAGCTCAGACTCTTATTTAGTAAATGGGACG 420
 DB 361 AGATCGTGGAGAAAGGGTTTAGGGTTTAAAGCTCAGACTCTTATTTAGTAAATGGGACG 420
 QY 421 GTGTCAATTTTCGGTTTGGAAATGAACCTTTGGGCTCAGCTTATGGGCTATTAGATATT 480
 DB 421 GTGTCAATTTTCGGTTTGGAAATGAACCTTTGGGCTCAGCTTATGGGCTATTAGATATT 480
 QY 481 TGATGGCTCTTCTAGTAAATACATATAGTTTATTTGGGCTTAGTTTAAATAGCCCATGT 540
 DB 481 TGATGGCTCTTCTAGTAAATACATATAGTTTATTTGGGCTTAGTTTAAATAGCCCATGT 540
 QY 541 TGGAAATATTTGACACATGCTTGGCTACTAGTCTAAACATGCAACCGAAGTGTGCG 600
 DB 541 TGGAAATATTTGACACATGCTTGGCTACTAGTCTAAACATGCAACCGAAGTGTGCG 600
 QY 601 AGACAAAGTCGAGCATATACAAATGGATCAAAACAGCCCTAGTGTGCGCGCTCTCGCTCAT 660
 DB 601 AGACAAAGTCGAGCATATACAAATGGATCAAAACAGCCCTAGTGTGCGCGCTCTCGCTCAT 660
 QY 661 GTGTCACTTTTCTCCTCGTTTCTTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTCA 720
 DB 661 GTGTCACTTTTCTCCTCGTTTCTTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTCA 720
 QY 721 ATACAAATTTTGGCTGATCTTCTGCAAACTCTTGCATATCGCCCAATATACGTGAACA 780
 DB 721 ATACAAATTTTGGCTGATCTTCTGCAAACTCTTGCATATCGCCCAATATACGTGAACA 780
 QY 781 CTGGTATCAATTTTGGTGTGTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAAG 840
 DB 781 CTGGTATCAATTTTGGTGTGTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAAG 840
 QY 841 TGAATTTATGATCATGTTTAAACATTTGATGATGATGATGATGATGATGATGATGATGATTT 900
 DB 841 TGAATTTATGATCATGTTTAAACATTTGATGATGATGATGATGATGATGATGATGATGATTT 900
 QY 901 AGTTGATGATTAACGTAAGCAAAATTAAGATAGATACATTTTGAATTTTGTGCTATTGTTG 960
 DB 901 AGTTGATGATTAACGTAAGCAAAATTAAGATAGATACATTTTGAATTTTGTGCTATTGTTG 960

DB 901 AGTTGATGATTAACGTAAGCAAAATTAAGATAGATACATTTGATTTTGTGCTATTGTTG 960
 QY 961 ACATATGCGAGAGTGAAGTACGCGCATGAAGATCAAGAGACACTTGTCTGAGCTCACAG 1020
 DB 961 ACATATGCGAGAGTGAAGTACGCGCATGAAGATCAAGAGACACTTGTCTGAGCTCACAG 1020
 QY 1021 AGTGAGCTGTAAAGAGCTTAGACTGAAGTCCCATGCAAACTTAATCCTACGTGGCTCAA 1080
 DB 1021 AGTGAGCTGTAAAGAGCTTAGACTGAAGTCCCATGCAAACTTAATCCTACGTGGCTCAA 1080
 QY 1081 ACCACGAGTCACTTGCACATATATAAACTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1140
 DB 1081 ACCACGAGTCACTTGCACATATATAAACTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1140
 QY 1141 CTCACAAACAAACAAAAG 1158
 DB 1141 CTCACAAACAAACAAAAG 1158
 RESULT 2
 ABZ17132
 ID ABZ17132 standard; DNA; 1223 BP.
 XX
 AC ABZ17132;
 XX
 DT 21-JAN-2003 (first entry)
 XX
 DE Arabidopsis thaliana stress regulated gene SEQ ID NO 4937.
 XX
 KW Arabidopsis thaliana; plant; gene; stress; transgenic; ds.
 XX
 OS Arabidopsis thaliana.
 XX
 PN WO200216655-A2.
 XX
 PD 28-FEB-2002.
 XX
 PF 24-AUG-2001; 2001WO-US026685.
 XX
 PR 24-AUG-2000; 2000US-0227866P.
 PR 26-JAN-2001; 2001US-0264647P.
 PR 22-JUN-2001; 2001US-0300111P.
 XX
 (SCRI) SCRIPPS RES INST.
 (SIGN) SYNGENTA PARTICIPATIONS AG.
 XX
 PI Harper JF, Kreps J, Wang X, Zhu T;
 XX
 WPI; 2002-304127/34.
 XX
 PT Identifying a stress condition to which a plant cell has been exposed and
 PT producing plants with increased tolerance to these abiotic stresses.
 XX
 PS Claim 14; SEQ ID NO 4937; 577pp + Sequence Listing; English.
 XX
 CC The invention relates to identifying a stress condition to which a plant
 CC cell has been exposed, comprising: (a) contacting nucleic acid
 CC representative of expressed polynucleotides in the plant cell with an
 CC array or probes representative of the plant cell genome; and (b)
 CC detecting a profile of expressed polynucleotides in the plant cell
 CC characteristic of a stress response. The method is useful in the
 CC production of transgenic plants, cells and seeds and in producing plants
 CC with increased tolerance to abiotic stress. The present sequence is that
 CC of an Arabidopsis thaliana stress regulated gene (ABZ12196-ABZ17574) used
 CC in methods of the invention. Note: The sequence data for this patent is
 CC not represented in the printed specification but is based on sequence
 CC information supplied to Derwent by the European Patent Office
 XX
 SQ Sequence 1223 BP; 421 A; 241 C; 212 G; 349 T; 0 U; 0 Other;
 Query Match 100.0%; Score 1158; DB 6; Length 1223;
 Best Local Similarity 100.0%; Pred. No. 4.6e-250;
 Matches 1158; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CACAAACATACACTCAAAATCCAGACTCACTCTACTCAATTTATGCACTTTCATCATGAA 60
Db |||||
QY 63 CACAAACATACACTCAAAATCCAGACTCACTCTACTCAATTTATGCACTTTCATCATGAA 122
Db |||||
QY 61 AACATCAAAACAGTCAAAAGTAAACAAATCAAGTCAGATTTCAGACACAAAGCCAGTAA 120
Db |||||
QY 123 AACATCAAAACAGTCAAAAGTAAACAAATCAAGTCAGATTTCAGACACAAAGCCAGTAA 182
Db |||||
QY 121 GATAGAAATTTAAGGACGCTCATCTGCTAAAGTGGGCAAAATCTTCTTAATCAAAACAG 180
Db |||||
QY 183 GATAGAAATTTAAGGACGCTCATCTGCTAAAGTGGGCAAAATCTTCTTAATCAAAACAG 242
Db |||||
QY 181 TAAACAGAGTAATTTAGCAAAATCCGAGCAGAAATCTCTACCCACCTCCGAAATTCACG 240
Db |||||
QY 243 TAAACAGAGTAATTTAGCAAAATCCGAGCAGAAATCTCTACCCACCTCCGAAATTCACG 302
Db |||||
QY 241 TCTTCACTAAATTTTGAAGGAATCGATCAATACCAACCCATTACACAAATACATAA 300
Db |||||
QY 303 TCTTCACTAAATTTTGAAGGAATCGATCAATACCAACCCATTACACAAATACATAA 362
Db |||||
QY 301 TCAAAATGGCGAGAACTGTAAGGAACTTGGCTTCAAGTCGAGAGAGGAAAGGA 360
Db |||||
QY 363 TCAAAATGGCGAGAACTGTAAGGAACTTGGCTTCAAGTCGAGAGAGGAAAGGA 422
Db |||||
QY 361 AGATCGTGGAGAAAGGGGTTTAGGGTTTAAAGCTCAGACTTCTATTGGAGTAAATGGGACG 420
Db |||||
QY 423 AGATCGTGGAGAAAGGGGTTTAGGGTTTAAAGCTCAGACTTCTATTGGAGTAAATGGGACG 482
Db |||||
QY 421 GTGTCACTTTCGGTTTTGAAATGAACCTTGGGCTCAGGTTATGGGCTATTAGATATT 480
Db |||||
QY 483 GTGTCACTTTCGGTTTTGAAATGAACCTTGGGCTCAGGTTATGGGCTATTAGATATT 542
Db |||||
QY 481 TGATGGCTTTCAGTAAATACAAATAGTATTAGGCTTAGTTTAAATGAAGCCCATGT 540
Db |||||
QY 543 TGATGGCTTTCAGTAAATACAAATAGTATTAGGCTTAGTTTAAATGAAGCCCATGT 602
Db |||||
QY 541 TGAATAATTTTGACATGCTTCTGGCTTACAGTGTAAACATCAACCGAACAGTGTGCG 600
Db |||||
QY 603 TGAATAATTTTGACATGCTTCTGGCTTACAGTGTAAACATCAACCGAACAGTGTGCG 662
Db |||||
QY 601 AGACATGCGAGATATACAAATGGATCAAAACGCGCTAGTGTGCGCGCTCTCGCTCAT 660
Db |||||
QY 663 AGACATGCGAGATATACAAATGGATCAAAACGCGCTAGTGTGCGCGCTCTCGCTCAT 722
Db |||||
QY 661 GTGTCACTTGTTCCTCGTTTTTTTTTAAATTTTCAAGTTCTTTTGTGTTTATCTTCA 720
Db |||||
QY 723 GTGTCACTTGTTCCTCGTTTTTTTTTAAATTTTCAAGTTCTTTTGTGTTTATCTTCA 782
Db |||||
QY 721 ATACAAATTTTGGCTGTATCTTGCAAACTCTTCGATCATATCGCAATATAGTGAACA 780
Db |||||
QY 783 ATACAAATTTTGGCTGTATCTTGCAAACTCTTCGATCATATCGCAATATAGTGAACA 842
Db |||||
QY 781 CTGGTATCTAAATTTGTTGTGTTAAATGTTAAATTTAGATTCTATTCTCGGTTTAAAG 840
Db |||||
QY 843 CTGGTATCTAAATTTGTTGTGTTAAATGTTAAATTTAGATTCTATTCTCGGTTTAAAG 902
Db |||||
QY 841 TGAATTTATATGATCATGCTTAAACATTTGTAAGTAAAGTATGATAATAAATGATAAATTT 900
Db |||||
QY 903 TGAATTTATATGATCATGCTTAAACATTTGTAAGTAAAGTATGATAATAAATGATAAATTT 962
Db |||||
QY 901 AGTTGATGATAACGTGAAGCAAAATAGAGTAGATACATTTGATTTTGTGTTTGTGTTG 960
Db |||||
QY 963 AGTTGATGATAACGTGAAGCAAAATAGAGTAGATACATTTGATTTTGTGTTTGTGTTG 1022
Db |||||
QY 961 ACATATGCGAGAGTGAAGTACGCGATGAAGTCAAGAGACACTTGTCTGAGCTCACAG 1020
Db |||||
QY 1023 ACATATGCGAGAGTGAAGTACGCGATGAAGTCAAGAGACACTTGTCTGAGCTCACAG 1082
Db |||||
QY 1021 AGTGACGTGTAAGAACTTAGACTGAAGTCCCATGCAAACTTAATCTTACGTGGCTCAA 1080
Db |||||
QY 1083 AGTGACGTGTAAGAACTTAGACTGAAGTCCCATGCAAACTTAATCTTACGTGGCTCAA 1142
Db |||||

QY 1081 ACCACGAGCTCAGCTTGCACATATATAAATCTCTCTAAGTCCCGTTCTTTCATCATCT 1140
Db |||||
QY 1143 ACCACGAGCTCAGCTTGCACATATATAAATCTCTCTAAGTCCCGTTCTTTCATCATCT 1202
Db |||||
QY 1141 CTCACAAACAAACAAAAG 1158
Db |||||
QY 1203 CTCACAAACAAACAAAAG 1220
Db |||||
RESULT 3
AAC33222/c
ID AAC33222 standard; DNA; 1118 BP.
XX AC AAC33222;
XX DT 17-OCT-2000 (first entry)
XX Arabidopsis thaliana DNA fragment SEQ ID NO: 2235.
XX Arabidopsis thaliana.
XX Hybridisation assay; genetic mapping; gene expression control;
XX protein identification; signal transduction pathway; metabolic pathway;
XX promoter; termination sequence; ss.
XX Arabidopsis thaliana.
XX PN EPI033405-A2.
XX XX
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-00301439.
XX PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 28-APR-1999; 99US-0130891P.
PR 30-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 04-MAY-1999; 99US-0132407P.
PR 05-MAY-1999; 99US-0132484P.
PR 06-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 07-MAY-1999; 99US-0132487P.
PR 11-MAY-1999; 99US-0132863P.
PR 14-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 18-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 20-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.

PR	14-JUN-1999;	99US-0139119P.	PR	11-AUG-1999;	99US-0148319P.
PR	16-JUN-1999;	99US-0139452P.	PR	12-AUG-1999;	99US-0148341P.
PR	16-JUN-1999;	99US-0139453P.	PR	13-AUG-1999;	99US-0148565P.
PR	17-JUN-1999;	99US-0139492P.	PR	13-AUG-1999;	99US-0148684P.
PR	18-JUN-1999;	99US-0139454P.	PR	16-AUG-1999;	99US-0149368P.
PR	18-JUN-1999;	99US-0139455P.	PR	17-AUG-1999;	99US-0149175P.
PR	18-JUN-1999;	99US-0139456P.	PR	18-AUG-1999;	99US-0149426P.
PR	18-JUN-1999;	99US-0139457P.	PR	20-AUG-1999;	99US-0149722P.
PR	18-JUN-1999;	99US-0139458P.	PR	20-AUG-1999;	99US-0149723P.
PR	18-JUN-1999;	99US-0139459P.	PR	20-AUG-1999;	99US-0149929P.
PR	18-JUN-1999;	99US-0139460P.	PR	23-AUG-1999;	99US-0149902P.
PR	18-JUN-1999;	99US-0139461P.	PR	23-AUG-1999;	99US-0149930P.
PR	18-JUN-1999;	99US-0139462P.	PR	25-AUG-1999;	99US-0150566P.
PR	18-JUN-1999;	99US-0139463P.	PR	26-AUG-1999;	99US-0150884P.
PR	21-JUN-1999;	99US-0139817P.	PR	27-AUG-1999;	99US-0151065P.
PR	22-JUN-1999;	99US-0139899P.	PR	27-AUG-1999;	99US-0151066P.
PR	23-JUN-1999;	99US-0140353P.	PR	30-AUG-1999;	99US-0151080P.
PR	23-JUN-1999;	99US-0140354P.	PR	31-AUG-1999;	99US-0151303P.
PR	24-JUN-1999;	99US-0140695P.	PR	31-AUG-1999;	99US-0151438P.
PR	28-JUN-1999;	99US-0140823P.	PR	07-SEP-1999;	99US-0151930P.
PR	30-JUN-1999;	99US-0141287P.	PR	10-SEP-1999;	99US-0152363P.
PR	01-JUL-1999;	99US-0141842P.	PR	13-SEP-1999;	99US-0153070P.
PR	02-JUL-1999;	99US-0142055P.	PR	15-SEP-1999;	99US-0153758P.
PR	06-JUL-1999;	99US-0142390P.	PR	16-SEP-1999;	99US-0154018P.
PR	08-JUL-1999;	99US-0142803P.	PR	20-SEP-1999;	99US-0154779P.
PR	09-JUL-1999;	99US-0142920P.	PR	22-SEP-1999;	99US-0155139P.
PR	12-JUL-1999;	99US-0142977P.	PR	23-SEP-1999;	99US-0155486P.
PR	13-JUL-1999;	99US-0143542P.	PR	24-SEP-1999;	99US-0155659P.
PR	14-JUL-1999;	99US-0143624P.	PR	28-SEP-1999;	99US-0156458P.
PR	15-JUL-1999;	99US-0144005P.	PR	29-SEP-1999;	99US-0156596P.
PR	16-JUL-1999;	99US-0144085P.	PR	04-OCT-1999;	99US-0157117P.
PR	16-JUL-1999;	99US-0144086P.	PR	05-OCT-1999;	99US-0157753P.
PR	19-JUL-1999;	99US-0144325P.	PR	06-OCT-1999;	99US-0157865P.
PR	19-JUL-1999;	99US-0144331P.	PR	07-OCT-1999;	99US-0158029P.
PR	19-JUL-1999;	99US-0144332P.	PR	08-OCT-1999;	99US-0158232P.
PR	19-JUL-1999;	99US-0144333P.	PR	12-OCT-1999;	99US-0158369P.
PR	19-JUL-1999;	99US-0144334P.	PR	13-OCT-1999;	99US-0159293P.
PR	19-JUL-1999;	99US-0144335P.	PR	13-OCT-1999;	99US-0159294P.
PR	20-JUL-1999;	99US-0144352P.	PR	14-OCT-1999;	99US-0159295P.
PR	20-JUL-1999;	99US-0144632P.	PR	14-OCT-1999;	99US-0159330P.
PR	20-JUL-1999;	99US-0144884P.	PR	14-OCT-1999;	99US-0159331P.
PR	21-JUL-1999;	99US-0144814P.	PR	14-OCT-1999;	99US-0159637P.
PR	21-JUL-1999;	99US-0145086P.	PR	14-OCT-1999;	99US-0159638P.
PR	21-JUL-1999;	99US-0145088P.	PR	18-OCT-1999;	99US-0159584P.
PR	22-JUL-1999;	99US-0145085P.	PR	21-OCT-1999;	99US-0160741P.
PR	22-JUL-1999;	99US-0145087P.	PR	21-OCT-1999;	99US-0160767P.
PR	22-JUL-1999;	99US-0145089P.	PR	21-OCT-1999;	99US-0160768P.
PR	22-JUL-1999;	99US-0145192P.	PR	21-OCT-1999;	99US-0160814P.
PR	23-JUL-1999;	99US-0145145P.	PR	21-OCT-1999;	99US-0160815P.
PR	23-JUL-1999;	99US-0145218P.	PR	22-OCT-1999;	99US-0160980P.
PR	23-JUL-1999;	99US-0145224P.	PR	22-OCT-1999;	99US-0160981P.
PR	26-JUL-1999;	99US-0145276P.	PR	22-OCT-1999;	99US-0160989P.
PR	27-JUL-1999;	99US-0145913P.	PR	25-OCT-1999;	99US-0161040P.
PR	27-JUL-1999;	99US-0145918P.	PR	25-OCT-1999;	99US-0161405P.
PR	27-JUL-1999;	99US-0145919P.	PR	25-OCT-1999;	99US-0161406P.
PR	28-JUL-1999;	99US-0145951P.	PR	25-OCT-1999;	99US-0161359P.
PR	02-AUG-1999;	99US-0146388P.	PR	26-OCT-1999;	99US-0161360P.
PR	02-AUG-1999;	99US-0146389P.	PR	26-OCT-1999;	99US-0161361P.
PR	03-AUG-1999;	99US-0147038P.	PR	28-OCT-1999;	99US-0161920P.
PR	04-AUG-1999;	99US-0147204P.	PR	28-OCT-1999;	99US-0161992P.
PR	05-AUG-1999;	99US-0147302P.	PR	28-OCT-1999;	99US-0161993P.
PR	05-AUG-1999;	99US-0147192P.	PR	29-OCT-1999;	99US-0162142P.
PR	06-AUG-1999;	99US-0147260P.			
PR	06-AUG-1999;	99US-0147303P.			
PR	06-AUG-1999;	99US-0147416P.			
PR	09-AUG-1999;	99US-0147493P.			
PR	09-AUG-1999;	99US-0147935P.			
PR	10-AUG-1999;	99US-0148171P.			

Query Match 7.0%; Score 81; DB 3; Length 1118;
Best Local Similarity 100.0%; Pred. No. 2.2e-08;
Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 322 CTGGAACCTTTGCTCAAGTCGCAGAGAGAGAGAGAGAGAGATCGTGAGAAAGGGGTTT 381
DB 81 CTGGAACCTTTGCTCAAGTCGCAGAGAGAGAGAGAGAGATCGTGAGAAAGGGGTTT 22

QY	392	AGGGTTTAAGCTCAGACTTCT	402
Db	21	AGGGTTTAAGCTCAGACTTCT	1
RESULT 4			
ABL34166/c			
ID	ABL34166	standard; DNA; 10326 BP.	
XX	AC	AC	
XX	ABL34166;		
DT	26-MAR-2002	(first entry)	
XX			
DE		Human immune system associated gene SEQ ID NO: 2139.	
XX			
KW	Human; Immune system disease; cytosine methylation; antiasthmatic;		
KW	antiarteriosclerotic; antilanaemic; cytostatic; nootropic;		
KW	neuroprotective; anti-HIV; anticonvulsant; ophthalmological;		
KW	antiinflammatory; antiarthritic; antidiabetic; antiporiatic;		
KW	acute myeloid leukaemia; cancer; eye disease; arteriosclerosis; anaemia;		
KW	neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;		
XX	ds.		
OS	Homo sapiens.		
XX			
PN	WO200200928-A2.		
XX			
PD	03-JAN-2002.		
XX			
PF	02-JUL-2001; 2001WO-EP007537.		
XX			
PR	30-JUN-2000; 2000DE-01032529.		
PR	01-SEP-2000; 2000DE-01043826.		
XX			
PA	(EPIG-) EPIGENOMICS AG.		
XX			
PI	Olek A, Piepenbrock C, Berlin K;		
XX			
DR	WPI; 2002-130909/17.		
XX			
PT	Nucleic acid comprising fragment of chemically modified gene, useful for		
PT	diagnosis and treatment of diseases associated with abnormal cytosine		
PT	methylation.		
XX			
PS	Claim 1; SEQ ID NO 2139; 32pp + Sequence Listing; German.		
XX			
CC	The present invention provides a number of human immune system associated		
CC	genes which are modified by the methylation of cytosines. The sequences		
CC	can be used in the diagnosis and treatment of immune system disorders,		
CC	including eye diseases such as retinopathy, neovascular glaucoma and		
CC	macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid		
CC	leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,		
CC	rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel		
CC	diseases. The present sequence is a gene of the invention		
XX			
SQ	Sequence 10326 BP; 2542 A; 104 C; 2689 G; 4991 T; 0 U; 0 Other;		
		5.6%; Score 65.4; DB 6; Length 10326;	
Query Match	Best Local Similarity	48.8%; Pred. No. 0.00011;	
Matches	177; Conservative	0; Mismatches 186; Indels 0; Gaps 0	
QY	2	ACAAACATACACTCAAATCCAGACTCACATCTACTCAATTATGCAACTTCATCATGAA	61
Db	2764	ACCAACTTAACCTCCAAAAAACCCTAAATTAATAACCTTAACAAAAATTCAACTAT	2705
QY	62	ACATCAAAACAGTCAAAGTAAACAAATCAAGTCAGATTCAGACACAAAGCCAGTAAAG	121
Db	2704	TCACCTCAAAACACAAAAATTTCTTACCTACTAAAAATCAACAAACAAACAAATTA	2645
QY	122	ATAGAAATTTAAGGAAGCGTCATGCTAAGCTGCGGAAATATCTTCTTAATCAAAACAGT	181

[illegible]

QY	76	CAAGTAACAAAAATCAAGTCAGATTCAGCACACAAACCGCAGTAAAGATGAAAAATTAAC	135
Db	985	TAAACAAAATAAAATCAAAAAAAAACCCCTCATACAATACTACTAAAAATATAAAATATTAC	926
QY	136	GAACGCTCATGCTAGCTCGCGAARATCTTCTTAATCAAAACAGTAAACACGAGTAATT	195
Db	925	AACCACTATAAAAAACAAATAAAAATCTCAAAAAATAAACTAAACCGATATATA	866
QY	196	AGCAAAATCGAGCAGAAAACTCTCAACCACCTCGAAATTCAGGTCCTCACTAAAAATT	255
Db	865	ATCTAACAAATCTTATAAATATATATCTTAAAAAAAATAAATCAATACATTTAAAAAATA	806
QY	256	TCGAAAGGAATCGATCAATACCAACCCATTACAAAAATACATAATCAAAA	306
Db	805	TCTACACTCCCATATTTATTACACACACTATTCACATTAACCTAAAAATATAA	755

RESULT 11	
ABL34526/c	
ID	ABL34526 standard; DNA; 6713 BP.
XX	
AC	ABL34526;
XX	
DT	26-MAR-2002 (first entry)
XX	
DE	Human metastasis associated gene SEQ ID NO: 79.
XX	
XX	Metastasis associated gene; cytostatic; gene therapy; cancer;
KW	cytosine methylation; gene; ds.
XX	
OS	Homo sapiens.
XX	
XX	WO200177376-A2.
XX	
PN	18-OCT-2001.
XX	
PD	
XX	
PF	06-APR-2001; 2001WO-EP003970.
XX	
PR	06-APR-2000; 2000DE-01019058.
XX	
PR	07-APR-2000; 2000DE-01019173.
XX	
PR	30-JUN-2000; 2000DE-01032529.
XX	
PR	01-SEP-2000; 2000DE-01043826.
XX	
XX	
PA	(EPIG-) EPIGENOMICS AG.
XX	
PI	Olek A, Piepenbrock C, Berlin K;
XX	
DR	WPI; 2002-010922/01.
XX	
PT	New nucleic acid derived from chemically treated metastasis genes, useful
PT	for diagnosis of cancers by analysis of cytosine methylation, also for
PT	treatment.
XX	
PS	Claim 1; SEQ ID NO 79; 23pp + Sequence Listing; English.
XX	
CC	The present invention provides a number of human metastasis associated
CC	genes which are modified by cytosine methylation. The sequences can be
CC	used in the diagnosis and treatment of cancer. The present sequence is
CC	one of the genes of the invention
XX	
SQ	Sequence 6713 BP; 1676 A; 149 C; 1441 G; 3447 T; 0 U; 0 Other;
Query Match	4.7%; Score 54.2; DB 6; Length 6713;
Best Local Similarity	49.1%; Pred. No. 0.033;
Matches 143;	Conservative 0; Mismatches 148; Indels 0; Gaps 0
QY	16 AAATTCGAGCTCAGATCTACTCAATTATGCACTTCATCATGAAACATCAAAAACAGT 75
	1045 AAAATATCACCTCACTCCCAATTTAAATAAATCTATATAAAAAAACAACCAATTACCA 986
Db	76 CAAAGTACAAAATCMAAGTCAGATTCAGCACACAAAGCCAGTAAAGATGAAAAATTAAC 135
QY	985 TACAAAATAAAATCAAAAAAAAACCCCTCATACAACTACTAAAAATATAAATTATTAC 926
Db	

136 GAACGCTCATGCTAAGCTCGCAAAATATCTTCTTAATCAAAAACAGTAAACACGAGTAATT 195
Qy
925 AACCACTATAAAAAACAAAATAAAATTTCTTCAAAAAATAAACTAAAAACGATATATA 966
Db
136 AGCAAAATCCGAGCAGAAAACTCTCCGCACTCCGAAATTCACGCTCTTTCATAAAAATTT 255
Qy
865 ATCTAACCAATTTCTTATAAATATATATCTTAAAAAATAAAATCAATACATTAAAAAATA 806
Db
256 TCGAAGAAGAAATCGATCAATACGAACCAATTACAAAAATACATAATCAAAA 306
Qy
805 TCTACACTCCCATATTTTATTACACACTATTTCATAATTAATCAAAAAATAAAA 755
Db

RESULT 12	
ABL70251/c	
ID	ABL70251 standard; DNA; 6713 BP.
XX	
AC	ABL70251;
XX	
DT	01-JUL-2002 (first entry)
XX	
DE	Chemically treated cell signalling DNA sequence#71.
XX	
KW	Cell signalling; cytosine methylation; cell signalling disease; cancer;
KW	tumour; cytostatic; ds.
XX	
OS	Unidentified.
XX	
FN	WO200202807-A2.
XX	
PD	10-JAN-2002.
XX	
PF	29-JUN-2001; 2001WO-EP007471.
XX	
PR	30-JUN-2000; 2000DE-01032529.
PR	01-SEP-2000; 2000DE-01043826.
XX	
PA	(EPIG-) EPIGENOMICS AG.
XX	
PI	Olek A, Piepenbrock C, Berlin K;
XX	
DR	WPI; 2002-154758/20.
XX	
PT	Nucleic acid, useful for diagnosis and therapy of diseases associated
PT	with cell signaling e.g. cancer, comprises chemically modified genomic
PT	sequences of genes associated with cell signaling.
XX	
PS	Claim 1; SEQ ID NO 141; 24pp + Sequence Listing; English.
XX	
CC	The invention relates to a nucleic acid comprising a sequence of at least
CC	18 bases of a segment of chemically pretreated DNA of genes associated
CC	with cell signalling. The activity of the modified sequences of the
CC	invention may be described as cytostatic. The object of the invention is
CC	to provide the chemically modified DNA of genes associated with cell
CC	signalling, as well as oligonucleotides and/or PNA-oligomers for
CC	detecting cytosine methylations, as well as a method which is
CC	particularly suitable for the diagnosis and/or therapy of genetic and
CC	epigenetic parameters of genes associated with cell signalling. The
CC	chemically modified DNA provided by the invention is useful for diagnosis
CC	and therapy of diseases such as solid tumours and cancer. The sequences
CC	given in records ABL70111-ABL70626 represent chemically pre-treated
CC	genomic DNA's of genes associated with cell signalling. Note: The
CC	sequence data for this patent is not represented in the printed
CC	specification, but is based on sequence information supplied by the
XX	European Patent Office
XX	
SQ	Sequence 6713 BP; 1676 A; 149 C; 1441 G; 3447 T; 0 U; 0 Other;

Query Match 4.7%; Score 54.2; DB 6; Length 6713;
Best Local Similarity 49.1%; Pred. NO. 0.033;
Matches 143; Conservative 0; Mismatches 148; Indels 0; Gaps 0;

CC (array) obtainable in (6); (8) an array of different oligonucleotide-
 CC and/or PNA-oligomer sequences, which are arranged on a plane solid phase
 CC in the form of a rectangular or hexagonal lattice; (9) a composition of
 CC matter comprising the nucleic acid and a buffer comprising 1-5 mM
 CC magnesium chloride, 100-500 micromole dNTP, 0.5-5 units of taq
 CC polymerase, and the oligomer; (10) detecting, differentiating or
 CC distinguishing between colon cell proliferative disorders; (11) detecting
 CC a colon cell proliferative disorder; and (12) a kit useful for the method
 CC in (10) comprising a bisulfite reagent, and at least one of the nucleic
 CC acid molecule or peptide described above, or their complements. EYA4 has
 CC cytosolic activity, and can be used in gene therapy. The methods, vector
 CC and polypeptide from the present invention are useful for treating colon
 CC cell proliferative disorders. The EYA4 gene, the polypeptide expressed
 CC from the EYA4 gene and kit are useful for detecting, differentiating or
 CC distinguishing between colon cell proliferative disorders. The oligomers
 CC are useful for detecting the cytosine methylation state and/or single
 CC nucleotide polymorphisms (SNPs) within nucleic acid sequences. The array
 CC is useful for analyzing diseases associated with the methylation state of
 CC the CpG dinucleotides. The present sequence represents a chemically
 CC pretreated human EYA4 gene, which is given in the exemplification of the
 CC present invention. Human EYA4 is mapped to chromosome 6, more
 CC specifically to 6q22.3.

Sequence 29993 BP; 8706 A; 0 C; 5856 G; 15431 T; 0 U; 0 Other;
 Query Match 4.6%; Score 53.4; DB 10; Length 29993;
 Best Local Similarity 47.7%; Pred. No. 0.07;
 Matches 156; Conservative 0; Mismatches 171; Indels 0; Gaps 0;

QY 4 AAACATACATCCAAATCCAGACTCAGATCTACTCTCAATTTGCACTTCATGAAAC 63
 DB 21003 AAAAAAAAAATATAAAACCATACACAAAACCAAAATCTAAATCTTAAACT 20944
 QY 64 ATCAAAACAGTCAAAAGTAAACAAATCAAGTCAAGTTCAGACACAAAGCCAGTAAGAT 123
 DB 20943 AAAAAAAAAACAAACAAACAAACAAACAAACAAATTCACCACTCTCTATAAAAAA 20884
 QY 124 AGAAATTTAAGAGCCTCATGCTAGCTGCGCAAAATCTCTTAATCAACAAACAGTAA 183
 DB 20883 AAAAAAAAAACAAATATCATCATCTACTCTAAATCTCTTATCAAAATCTAACTAA 20824
 QY 184 CAAGAGTAATTAGCAAAATCCAGCGAGAAATCTCCACCACTCGGAAATTCAGTCT 243
 DB 20823 AAAAAATTAACAACTTTCTCTCAITTCAAATTTACTTATAACCCCAACAAATCAATCA 20764
 QY 244 TCACTAAAATTTTCGAAAGGAATCGATCAATACCAACCCCAATTCACACAAATCAATCA 303
 DB 20763 AAAAAATTAACCAAACTTAATAAAAAACAAATTTAATAAAAAACACAAAAA 20704
 QY 304 AATGCGAGATCGTACCTGGAATCT 330
 DB 20703 TACTAATTAACCTAATAAAAAAAAT 20677

RESULT 15
 ADA71938
 ID ADA71938 standard; DNA; 2000 BP.
 AC ADA71938;
 DT 20-NOV-2003 (first entry)
 DE Rice gene, SEQ ID 5263.
 XX Plant; bacterial infection; fungal infection; viral infection; rice;
 KW gene; ds.
 XX Oryza sativa.
 OS
 XX WO200300898-A1.
 PN
 XX 03-JAN-2003.
 PD
 XX

PF 22-JUN-2001; 2001WO-IB001105.
 XX
 PR 22-JUN-2001; 2001WO-IB001105.
 XX
 PA (SYGN) SYNGENTA PARTICIPATIONS AG.
 XX
 PI Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y,
 PI Katagiri F, Quan S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;
 XX WPI; 2003-175290/17.
 XX
 PT Identifying at least one gene involved in plant resistance or response to
 PT pathogenic infection for conferring resistance or tolerance to a plant to
 PT bacterial, fungal or viral infection by determining or detecting plant
 XX gene expression.
 PS Claim 27; SEQ ID NO 5263; 899pp; English.
 XX
 CC The present invention relates to a method (M1) for identifying genes
 CC involved in plant resistance or response to pathogenic infection. M1
 CC comprises identifying a gene whose expression is significantly altered in
 CC the incompatible interaction of plant gene expression relative to
 CC expression of the gene in an uninfected plant, in a mutant plant that
 CC does not express a gene associated with response to pathogenic infection,
 CC or in a corresponding incompatible or compatible interaction. (M1) is
 CC useful for conferring resistance or tolerance to a plant to
 CC bacterial, fungal or viral infection. The present sequence was used to
 CC illustrate the invention.
 XX
 SQ Sequence 2000 BP; 336 A; 265 C; 284 G; 363 T; 0 U; 752 Other;
 Query Match 4.5%; Score 52.4; DB 8; Length 2000;
 Best Local Similarity 8.9%; Pred. No. 0.065;
 Matches 73; Conservative 375; Mismatches 361; Indels 9; Gaps 2;

QY 230 CGAAATTCAGTCTTCACTAAATTTTCGAAAGAAATCGATCAATACCAACCAATACAC 289
 DB 188 SGSMKRMNMSSCGRCGCRSSAYSYGYTGRKYGYKMYYSASRCMEAYMTTSWSA 247
 QY 290 AAAATACATAATCAAAATGCGAGATCGTACCTGGAACCTTTGCTCAAGTCGACAGA 349
 DB 248 CSYTWCRSKRSMKRMKMSYGYMSYKMYWYRGYKMGWAGRMMSKACACACACAC 307
 QY 350 GAGAAAAAGAGATCGTGGAGAAAGGGTGTAGGGTTTAAAGCTCAGACTCTATTGGAG 409
 DB 308 RGATRYWGRYMSRMAMMYKMYWYRGYKMGWAGRMMSKACACACACACACAC 367
 QY 410 TAAATGGGACGGTGTCACATTTTCGGTTTGGAAATGAACTTTGGGCTCACGTTATGGC 469
 DB 368 RMTRRRWAKSRSTSRKKKMKCRKRYKMRGYSRMRSCRAARWKRCSGRWKG 427
 QY 470 TATTAGATATTTCATGGGCTTTCTAGTAAATACATAATAGTTATTTGGCTTAGTTAA 529
 DB 428 CRGCTCMKSYGMRKSKKMSKMSYKMYWYRGYKMGWAGRMMSKACACACACAC 479
 QY 530 TAAAGCCATGTTGAAATATTTCACACATGCTTGGCTACTAGTCTAAACATCAAC 589
 DB 480 GMMGTGCRVYKRSKMRKCRRRRWGRMYRMYRMYRMYRMYRMYRMYRMYRMYRMYR 539
 QY 590 AACAGTTGTCGAGACAAAGTCGACGATATACATGATCAACACGCTAGTGTGCGCGC 649
 DB 540 ARCWTYRGYWGMMKRYKMYWYKMYWYKMYWYKMYWYKMYWYKMYWYKMYWYKMYWY 599
 QY 650 GTCTCGCTCATGTGTCACCTTTCTCTCGTTTCTCTTTTAAATTTTCAATGTTCTTTG 709
 DB 600 KWKRSKMSKMSKMSKMSKMSKMSKMSKMSKMSKMSKMSKMSKMSKMSKMSKMSKMS 659
 QY 710 TTTTATCTCAATACAAATTTTGGCTGTATCTTCAAACTCTTCGATCATATCCCAAT 769
 DB 660 SMKSTCTWYMYKMYKMYWYRMYRMYRMYRMYRMYRMYRMYRMYRMYRMYRMYRMYR 719
 QY 770 ATAGTGAACTCGTATCAATTTGTGTGTTAATTTGTTAAATTTAGATTCTATCTC 829
 DB 770 ATAGTGAACTCGTATCAATTTGTGTGTTAATTTGTTAAATTTAGATTCTATCTC 829

Thu Oct 14 09:37:08 2004

Db 720 MTGMYSGRYWTSWKYKCKSWKYSRMWYWSWWWAKTWKMMRRYATRMMWYRYSMKW 779
QY 830 CGGTTTAAAGTGAATTATATGTCATCGTTTAAAAACATTGTAAAGTAAGATGATAATAAA 889
Db 780 YTWCTWGYWYWWRTYMKRMYWYKCTKYWYWSATYWTGTWAAWWWAKTKMRMGMTGA 839
QY 890 ATGATAAATTTTACTTGATGATAACGTGAAGCAAAAAATGAGATAGATACATTGATTTT 949
Db 840 KTFGRARAKARYWVKWATWCATKRWMTKGKAKWATWMAKWRKYYSWNRWYKYKTR 899
QY 950 GTCGTATTTTGAACATATGCGGAGAGTGAGCTACGCGCATCAAGATCAAGAGACACTTGCT 1009
Db 900 RT-RYKTCWVKARWGSWAYWRMWWKSGAKWMMWKKGRWGTKYWYCTTWKMACGRAT 958
QY 1010 CGAGCTCACAGAGTGACGTGTAAAGCTTTAGACTGAA 1047
Db 959 KYMCCAGWNAWYSYSWRTYWMRTWRWMASSRTAKEM 996

Search completed: October 9, 2004, 04:43:59
Job time : 870 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2004, 03:54:36 ; Search time 145 Seconds
(without alignments)
5676.507 Million cell updates/sec

Title: US-09-998-059-1
Perfect score: 1158
Sequence: 1 cacaacatacactcaaat.....ctctcacaacacacaaaaag 1158

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 824507 seqs, 355394441 residues

Total number of hits satisfying chosen parameters: 1649014

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued patents NA:*

- 1: /cgn2_6/ptodata/1/ina/5A COMB.seq.*
- 2: /cgn2_6/ptodata/1/ina/5B COMB.seq.*
- 3: /cgn2_6/ptodata/1/ina/6A COMB.seq.*
- 4: /cgn2_6/ptodata/1/ina/6B COMB.seq.*
- 5: /cgn2_6/ptodata/1/ina/PTUS COMB.seq.*
- 6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	53.6	4.6	7215	3	US-09-134-001C-627
2	52	4.5	694	4	US-09-710-279-1601
3	52	4.5	3028	4	US-09-710-279-3860
4	52	4.5	3032	4	US-09-710-279-4227
5	52	4.5	6968	4	US-09-710-279-759
6	49.8	4.3	4140	3	US-08-894-731-2
7	48.8	4.2	1851	4	US-09-601-198-51
8	48.6	4.2	1141	4	US-09-806-708B-22
9	47.4	4.1	1141	4	US-09-806-708B-22
10	45.8	4.0	7218	1	US-09-806-708B-22
11	45.6	3.9	8607	4	US-10-204-708-71
12	45.4	3.9	11049	4	US-10-204-708-23
13	44.4	3.8	6801	4	US-10-204-708-62
14	43.6	3.8	3600	3	US-08-894-731-1
15	43.2	3.7	1316	4	US-09-270-767-2481
16	43.2	3.7	1316	4	US-09-270-767-17763
17	43	3.7	6156	4	US-10-204-708-59
18	42.8	3.7	640681	4	US-10-204-708-8
19	42.4	3.7	6020	4	US-10-204-708-77
20	42.2	3.6	5915	4	US-10-204-708-77
21	42	3.6	66933	4	US-09-544-398B-11
22	42	3.6	66933	4	US-09-543-771-11
23	42	3.6	72049	4	US-09-544-398B-9
24	42	3.6	72049	4	US-09-543-771-9
25	41.8	3.6	519	4	US-09-248-796A-6405
26	41.4	3.6	658	3	US-08-998-416-595
27	41.4	3.6	51952	3	US-08-947-823-1

C 28	41.2	3.6	10144	4	US-10-204-708-94	Sequence 94, Appl
C 29	41.2	3.6	10467	4	US-10-204-708-2	Sequence 2, Appl
C 30	41	3.5	8607	4	US-10-204-708-72	Sequence 72, Appl
C 31	40.6	3.5	5562	4	US-10-204-708-63	Sequence 63, Appl
C 32	40.6	3.5	6156	4	US-10-204-708-60	Sequence 60, Appl
C 33	40.6	3.5	7218	1	US-08-232-463-14	Sequence 14, Appl
C 34	40.6	3.5	7786	4	US-09-790-988-2	Sequence 2, Appl
C 35	40.6	3.5	8607	4	US-10-204-708-72	Sequence 72, Appl
C 36	40.6	3.5	11049	4	US-10-204-708-23	Sequence 23, Appl
C 37	40.4	3.5	600	4	US-09-601-198-142	Sequence 142, App
C 38	40.4	3.5	1103	3	US-09-122-400B-20	Sequence 20, Appl
C 39	40.4	3.5	1167	1	US-08-592-658-1	Sequence 1, Appl
C 40	40.4	3.5	2672	1	US-08-703-947-1	Sequence 1, Appl
C 41	40.4	3.5	6040	4	US-10-204-708-69	Sequence 69, Appl
C 42	40.2	3.5	6306	4	US-10-204-708-50	Sequence 50, Appl
C 43	40.2	3.5	640681	4	US-09-790-988-1	Sequence 1, Appl
C 44	40.2	3.5	786431	4	US-09-751-389-3	Sequence 3, Appl
C 45	40	3.5	4832	4	US-09-457-037B-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1
US-09-134-001C-627
; Sequence 627, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-08-14
; PRIOR APPLICATION NUMBER: US 60/055,779
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 627
; LENGTH: 7215
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-627

Query Match	4.6%	Score 53.6;	DB 3;	Length 7215;
Best Local Similarity	52.2%;	Pred. No. 0.00061;		
Matches 119;	Conservative 0;	Mismatches 109;	Indels 0;	Gaps 0;
QY	697	ATAAGTCTTTTGGTTTATCTTCAATACAAAATTTTGGCTGTATCTTGCAAACTCTTCA	756	
DB	372	ATTACCTCTGTTAATTTAAGATTATCAATTTCTGATGAAGCTACTCTTTTGA	431	
QY	757	TCATATCCCATATACGTGAACACTGGTATCTAATTTGTTGTTAATTTGTTAAATTT	816	
DB	432	TTTCACTCTACTATTTTAAATGAAGTCAATTTTCAACAAATTAATAT	491	
QY	817	AGATTCTATTCTCCGGTTTAAAGTGAATATATGATCATCGTTAAACATTTGAAGTA	876	
DB	492	CGCTTATACCACTCTTTTAAAGTGAATATAGTATCTCGGATTTTATATGATGATAA	551	
QY	877	AGATGATTAATAATGATATAATTTAGTTGATGATGAACGTAAGCAAA	924	
DB	552	AGAAATGAATCAAAAGATGATCAAAATCAAGTTGTGTGTCATCAAA	599	

RESULT 2
US-09-710-279-1601
; Sequence 1601, Application US/09710279
; Patent No. 6703492
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS

Db	2700	TGCTTATACCCACTCTTTTAAAGAGTAATAGTATCTCGGATTTTATAATTTGGATTAA	2759
Qy	877	AGATGATAATAAAATGATAAATTTAGTTGATGGATAACGTGAAGCAA	924
Db	2760	AGAAATGAATCAAAAGATGGATCAAAATCAAGTTGTGTCCATCAA	2807
RESULT 4			
US-09-710-279-4227			
; Sequence 4227, Application US/09710279			
; Patent No. 6703492			
; GENERAL INFORMATION:			
; APPLICANT: KIMMERLY, WILLIAM JOHN			
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS			
; FILE REFERENCE: PU34800S			
; CURRENT APPLICATION NUMBER: US/09/710,279			
; CURRENT FILING DATE: 2000-11-09			
; PRIOR APPLICATION NUMBER: 60/164,258			
; PRIOR FILING DATE: 1999-11-09			
; NUMBER OF SEQ ID NOS: 4472			
; SOFTWARE: PatentIn Ver. 2.1			
; SEQ ID NO 4227			
; LENGTH: 3032			
; TYPE: DNA			
; ORGANISM: Artificial Sequence			
; FEATURE:			
; OTHER INFORMATION: Description of Artificial Sequence: synthetic			
; OTHER INFORMATION: nucleic acid sequence			
US-09-710-279-4227			
Query Match			
Best Local Similarity			
Matches 118; Conservative			
Score 52; DB 4; Length 3032;			
Pred. No. 0.0011;			
Mismatches 110; Indels 0; Gaps 0;			
Qy	697	ATAAGTTCTTTTCTTTTATCTTCAATACAAAATTTTGGCTGTATCTTGCAAACTCTTCGA	756
Db	2689	ATTACCTCTGTTTAAATTTTAAAGATATATCAATTTCTAGATGGCGCTACCTACTTTTGA	2748
Qy	757	TCATATCGCAATATAGTGAACACTGGTGATCTAAATTTGTGTGTTAAATTTTAAATTT	816
Db	2749	TTTCCAGCTACTATTTTAAATGAAGTCAATTAACCTCCATTTTACACAAATTAATAT	2808
Qy	817	AGATTCATCTCCGGTTTAAAGTGAATATATATATATATATATATATATATATATATAT	876
Db	2809	TGCTTATACCCACTCTTTTAAAGTGAATATATATATATATATATATATATATATATAT	2868
Qy	877	AGATGATAATAAAATGATAAATTTAGTTGATGGATAACGTGAAGCAA	924
Db	2869	AGAAATGAATCAAAAGATGGATCAAAATCAAGTTGTGTGTCCATCAA	2916
RESULT 5			
US-09-710-279-759			
; Sequence 759, Application US/09710279			
; Patent No. 6703492			
; GENERAL INFORMATION:			
; APPLICANT: KIMMERLY, WILLIAM JOHN			
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS			
; FILE REFERENCE: PU34800S			
; CURRENT APPLICATION NUMBER: US/09/710,279			
; CURRENT FILING DATE: 2000-11-09			
; PRIOR APPLICATION NUMBER: 60/164,258			
; PRIOR FILING DATE: 1999-11-09			
; NUMBER OF SEQ ID NOS: 4472			
; SOFTWARE: PatentIn Ver. 2.1			
; SEQ ID NO 759			
; LENGTH: 6968			
; TYPE: DNA			
; ORGANISM: Artificial Sequence			
; FEATURE:			
; OTHER INFORMATION: Description of Artificial Sequence: synthetic			
; OTHER INFORMATION: nucleic acid sequence			
US-09-710-279-759			
Query Match			
Best Local Similarity			
Matches 118; Conservative			
Score 52; DB 4; Length 3028;			
Pred. No. 0.0011;			
Mismatches 110; Indels 0; Gaps 0;			
Qy	697	ATAAGTTCTTTTCTTTTATCTTCAATACAAAATTTTGGCTGTATCTTGCAAACTCTTCGA	756
Db	2580	ATTACCTCTGTTTAAATTTTAAAGATATATCAATTTCTAGATGGCGCTACCTACTTTTGA	2639
Qy	757	TCATATCGCAATATAGTGAACACTGGTGATCTAAATTTGTGTGTTAAATTTTAAATTT	816
Db	2640	TTTCCAGCTACTATTTTAAAGTGAATATATATATATATATATATATATATATATATAT	2699
Qy	817	AGATTCATCTCCGGTTTAAAGTGAATATATATATATATATATATATATATATATATAT	876

Search completed: October 9, 2004, 08:40:08
Job time : 148 secs

```
QY 658 CATGTGTCACCTGTGTTCTCGTCTTTTAAATTTTCAATAAGTCTTTTGTGTTTACT 717
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2555 CATTTCAAAATCGAATTCAGTAATATATATCTATCTATCGAATCGAAATCTTTCCAT 2496
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 718 TCAATACAAATTTTGGCTGTATCTGCAAACTCTCGAATCATATCGCAATATACG--T 775
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2495 TTAATAAAGTGAATGTGTGAATTTTTCATTTTAAATGTGAATGTGCGAAG 2436
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 776 GAACACTGGTGATCTAAATTTCTGTGTTAAATTTGTTAAATTTAGATTTCTCGGTTT 835
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2435 TTCCAAGAGAGTCTAGAAATTTCTTTTGTGCTTTTATTTACATTTGCTAGGTG 2376
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 836 AAAAGTGAATATATGATATCATGTTTAAACATTTGAAGTAAGATGATAATAAATGATA 895
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2375 ATTAAGAGATTTAAATTTATATTTATGAGTTTATTTGAAATTAATTATATTTTA 2316
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 896 AATTAGTTGATGATAACGTAAGCAAAATGAGATAGATACATTTGATTTTGTGCTA 955
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2315 AGAGTAGTTTGAAATAACTAAAGAGTAAATAAATAAATAATTAATTATATATCTA 2256
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 956 TTTT 959
||| |||
Db 2255 ATTT 2252
||| |||

RESULT 15
US-09-270-767-2481
; Sequence 2481, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2481
; LENGTH: 1316
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
; FEATURE:
; OTHER INFORMATION: n means any nucleotide
US-09-270-767-2481

Query Match 3.7%; Score 43.2; DB 4; Length 1316;
Best Local Similarity 46.1%; Pred. No. 0.14;
Matches 143; Conservative 0; Mismatches 163; Indels 4; Gaps 1;

QY 672 TTTCCTCGTTTTTTTAAATTTTTCATAAGTTCTTTTGTGTTTATCTTCAATACAAATTT 731
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 143 TTTTNTNNNTATTTTNTNNTAAATAAATATTTNGGGGTTATAAAATATAAAAAATA 202
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 732 TGGCTGTATCTTGCACACTCTTCGATCATATCGCAATATAGTGAACACTCGTGATCTA 791
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 203 TTTTATTTTAAATAATTTTATNTAATAATANNNAATAAATAAANTTTTTTTT 262
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 792 ATTTGTGTGTTAATG---TTAAATTTAGATTTCTTCTCGGTTTAAAGTGAATTA 847
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 263 ATATATTGGTATATGGAATTAATANTATATTTAATAAATTAATAAANTTAATTA 322
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 848 TATGTATCATGTTTAAACAACTTGAAGTAAAGATGATAATAAATGATAAATTTAGTTGAT 907
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 323 AATTTTATGTTNAAAAATNTAAATATTAATAATAAATAAATAAATAAATTTAT 382
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 908 GGATAACGTGAAGCAAAATGAGATAGATACATTTGATTTTGTGCTATTTTGACATATG 967
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 383 ATATGGAGATAAATAATTTTATGGGGAATTTNTGTAAATGTTTTTTTTTTTATTGAAGA 442
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 968 CGGAGAGTGA 977
||| ||| |||
Db 443 GTAGTGGGA 452
||| ||| |||
```

